THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED BY H. C. CARVER

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Volume X

1939

CONTENTS OF VOLUME X

Articles

NOTES

MISCELLANEOUS

A CONTRIBUTION TO THE THEORY OF SELF-RENEWING AGGREGATES, WITH SPECIAL REFERENCE TO INDUSTRIAL REPLACEMENT

By Alfred J. Lotka

1. **Introduction.** The analysis of problems of industrial replacement forms part of the more general analysis of problems presented by "self-renewing aggregates."[1] While the subject could, therefore, be treated in general and consequently rather abstract terms, for the purpose of exposition it will be advantageous to relate the discussion to concrete applications. These, in the past, have been mainly of two kinds, namely, first, applications to population analysis with related problems in genetics on the one hand and actuarial problems on the other; and second, applications to industrial replacement. As the fundamental setting of the two types of problems is very similar, leading in each case to certain integral equations, it will be advantageous to consider together both problems, or both phases of the general problem. This will incidentally give us an opportunity to observe the analogy, but also certain points of difference, between the two aspects of the problem.

Historically, the investigation of an actuarial problem came first. L. Herbelot[2] (1909) examined the number of annual accessions required to maintain a body of N policyholders constant, as members drop out by death. He assumes an initial body of N "charter" members at time $t = 0$, all of the same age, which for simplicity may be called age zero, since this merely amounts to fixing an arbitrary origin of the age scale. He further assumes the same uniform age at entry for each "new" member.

Then, if $p(t)$ is the probability at the age of entry of surviving t years, the survivors of charter members at time t will number $Np(t)$; and if $f(\tau)$ is the rate per head at which members drop out by death at time τ, being then immediately replaced by a new member of the fixed age of entry, then the survivors at time t of "new" members will evidently be given by

$$N \int_0^t f(\tau)p(t-\tau)\,d\tau$$

[1] I use here an English equivalent, as nearly as possible, to the German phrase "sich erneuernde Gesamtheiten," used by Swiss actuaries.

[2] Herbelot's original paper is disfigured with a number of misprints. It is essentially reproduced, with the errors corrected, in a paper by R. Risser (1912). The same treatment of the problem is also given by Zwinggi (1931) and by Schulthess (1935), (1937).

Hence, the condition for a constant membership N is

$$Np(t) + N\int_0^t f(\tau)p(t-\tau)\,d\tau = N \tag{1}$$

or

$$p(t) + \int_0^t f(\tau)p(t-\tau)\,d\tau = 1 \tag{2}$$

Differentiating with regard to t, and remembering that $p(0) = 1$, we have

$$p'(t) + \int_0^t f(\tau)p'(t-\tau)\,d\tau + f(t) = 0 \tag{3}$$

Equation (3) may be written

$$f(t) = -p'(t) - \int_0^t f(\tau)p'(t-\tau)\,d\tau \tag{4}$$

or, putting $(t-\tau) = a$

$$f(t) = -p'(t) - \int_0^t f(t-a)p'(a)\,da. \tag{5}$$

For the solution of the integral equation thus obtained Herbelot uses the method of successive differentiations,[3] duly pointing out its limitations, and applying it to several specific expressions for the survival function $p(a)$.

There is nothing in Herbelot's treatment to limit its application to living organisms. It is directly applicable to the problem of industrial replacement of an equipment comprising N original units installed at time $t = 0$, and maintained constant by the replacement of disused units with new.

Next in chronological order, of publications dealing with the type of problem with which we are here concerned, is a paper by Sharpe and Lotka (1911), who use Hertz's form of solution for the integral equation involved.[4] To this I wish

[3] This method is also followed in dealing with the problem of renewal by Risser (1912), (1920); Zwinggi (1931); Schulthess (1935), (1937); Preinreich (1938). All these authors applied their reflections to arbitrarily assumed frequency distributions for the renewal function, of simple analytical form. For example, among the more recent applications is one by Schulthess, who uses the function $p(t) = \left(1 - \frac{t}{\omega}\right)^m$; and quite recently, Preinreich has suggested the use of a Type I Pearson frequency curve on the basis of Kurtz's observational data. It is to be noted, however, that when it comes to actual application, Preinreich does not use an ordinary Pearson Type I curve nor actual observational data of any kind, but very conveniently simplifies the Pearson formula by giving integral values, namely 1 and 2, to the exponents, thereby reducing to triviality the task of applying the method of differentiation. None of these authors makes any attempt to deal with actual numerical observations which, in practice, fall far wide of any of the simple analytical formulae employed by them.

[4] P. Hertz, *Mathematische Annalen*, 1908, vol. 65, pp. 84 to 86.

to refer in some detail, adding to the original exposition in the light of later developments. The treatment of the subject proceeds here along somewhat broader lines, but, with obvious changes in the meaning of the symbols, and with certain modifications and limitations which are themselves of interest, the development is immediately applicable to economic systems composed of units having a characteristic "mortality" in use.

A population of living organisms, unlike industrial equipment, has practically no beginning. We know its existence only as a continuing process. Accordingly the equation for its development is most naturally framed without *explicit* reference to any "charter members."

The basis of the analysis is as follows:

In a population growing solely by excess of births over deaths (i.e. in the absence of immigration and emigration), the annual female births $B(t)$ at time t are the daughters of mothers a years old, born at time $(t - a)$ when the annual female births were $B(t - a)$. If fertility and mortality are constant and such that a fraction $p(a)$ of all births survive to age a, and are then reproducing at an average rate $m(a)$ daughters per head per annum, then, evidently,[5]

$$B(t) = \int_0^\infty B(t - a)p(a)m(a)\,da \tag{6}$$

$$= \int_0^\infty B(t - a)\varphi(a)\,da. \tag{7}$$

This is the fundamental equation in its original form, and, as noted above, it does not explicitly refer to any initial state, though, as will be seen presently, in order to make the problem determinate, data regarding the system at some particular period must be given. For the present we note that (7) can be written

$$B(t) = \int_t^\infty B(t - a)\varphi(a)\,da + \int_0^t B(t - a)\varphi(a)\,da \tag{8}$$

$$B(t) = B_1(t) + \int_0^t B(t - a)\varphi(a)\,da. \tag{9}$$

It is to be noted that the right hand member of (8), splits the total births $B(t)$ into two sections, those in which $(t - a) < 0$, that is, births of daughters whose mothers were born *before* $t = 0$; and those for which $(t - a) > 0$, that is births of daughters whose mothers were both *after* $t = 0$. The former section is denoted by $B_1(t)$ in (9). The function $B_1(t)$ thus defined will be found, in the

[5] Here and elsewhere in these developments the limits of the integral have, for simplicity, been written 0 and ∞. This ensures the inclusion of all nonvanishing terms in the integrand; the inclusion of terms for which either $\varphi(a)$ or $B(t - a)$ vanishes does not, of course, affect the value of the integral. If $\varphi(a)$ is represented between the limits α, ω of the reproductive period by some analytical expression, such as a Pearson frequency function, it is, of course, understood that outside the range α, ω we must put $\varphi(a) = 0$

further development, to play a significant rôle. Here it will suffice to point out that it vanishes for all values of t greater than ω, the upper limit of the reproductive period, because $\varphi(a)$ vanishes for these values of a.

2. **Special case.** A case of special interest is that in which $B_1(t)$ represents the births of daughters whose mothers were all born in an interval of time $t = -dt$ to $t = 0$. In that case the first integral in (8) reduces to a single term, so that

$$B(t) = B(0)\varphi(t)\,dt + \int_0^t B(t-a)\varphi(a)\,da \tag{10}$$

or, putting

$$B(0)\,dt = N_0 \tag{11}$$

$$B(t) = N_0\varphi(t) + \int_0^t B(t-a)\varphi(a)\,da. \tag{12}$$

This last equation holds also if a finite number of births take place (or are regarded as taking place) at a point of time $t = 0$.

Equations (10) and (12) are of interest as basic for the examination of the progeny of an infinitesimal population element,[6] that is, of a "zero" generation, born at time zero. In that case $B_1(t)$ is the annual rate of births in the "first" generation, and is simply proportional to $\varphi(t)$, i.e.

$$B_1(t) = N_0\varphi(t) \tag{13}$$

For the sake of greater generality the development has so far been given in terms of the phenomenon of replacement (reproduction) as it presents itself in a population of living organisms. But it should be noted here that, with appropriate changes in the meaning of $\varphi(a)$, equation (12) is directly applicable to the problem of industrial renewal in an installation originally installed at some point of time and maintained at a constant level by the replacement of each unit by a new one, the moment it is disused. In that case the "rate per head of reproduction" $m(a)$ at age a is evidently the same thing as the "death rate per head" at age a, namely

$$\mu(a) = -\frac{dp(a)}{p(a)\,da} = -\frac{p'(a)}{p(a)} \tag{14}$$

so that

$$\varphi(a) = p(a)\mu(a) \tag{15}$$

becomes

$$\varphi(a) = -p'(a). \tag{16}$$

[6] A. J. Lotka, (1928), (1929).

Reverting now to the fundamental equation in its first form (6), a trial substitution

$$B(t) = Qe^{rt} \tag{17}$$

is found to satisfy this equation, provided that r is a root of the characteristic equation

$$\int_0^\infty e^{-ra}\varphi(a)\,da = 1 \tag{18}$$

We may speak of (17) as a particular solution of (6) or (7). It is easily seen that the sum of such particular solutions is also a solution, i.e.

$$B(t) = Q_1e^{r_1t} + Q_2e^{r_2t} + \cdots \tag{19}$$

where r_1, r_2 etc., are roots of the characteristic equation (18).[7]

For real values of r the function

$$F(r) = \int_0^\infty e^{-ra}\varphi(a)\,da \tag{20}$$

decreases monotonically as r increases, since, from its nature, $\varphi(a) > 0$ for all values of a. Hence (18) can have only one real root r_1, and we shall have

$$r_1 \gtreqless 0 \quad \text{according as} \quad \int_0^\infty \varphi(a)\,da \gtreqless 1. \tag{21}$$

If $u + iv$ is a complex root of (18) then

$$1 = \int_0^\infty e^{-ua}\cos va\ \varphi(a)\,da \tag{22}$$

$$0 = \int_0^\infty e^{-ua}\sin va\ \varphi(a)\,da \tag{23}$$

and it is evident from (22) that $u < r_1$, since $\cos(va) \leqq 1$ for all values of a. The real part of any complex root of (18) is, therefore, algebraically less than the real root r_1.

This reasoning[8] is evidently quite independent of the particular form of $\varphi(a)$, and is thus equally true, whether $\varphi(a)$ be given in purely empirical form (defined by a table of values), or as a standard form of frequency curve, such as for example a Pearson curve of suitable type.

The roots of (18) can be determined directly, though rather laboriously, from

[7] For a discussion of the convergence of the series (19) see G. Herglotz, *Mathem. Annalen*, 1908, vol. 65, pp. 87 et seq.

[8] Adapted from P. Hertz, *Math. Annalen*, 1908, vol. 65, pp. 1–86; G. Herglotz, ibid. pp. 87–106. The Hertz solution is also applied to a similar problem by J. B. S. Haldane, *Proc. Cambridge Phil. Soc.*, 1926, vol. 23, p. 607. A particularly detailed development is given by H. T. J. Norton, *Proc. London Math. Soc.*, 1926, vol 28, p. 21.

equations (22) and (23); or, they can be brought into relation with the Thiele semivariants μ of the function $\varphi(a)$ defined by

$$F(r) = \int_0^\infty e^{-ra}\varphi(a)\,da = m_0 e^{-\mu_1 r + \frac{1}{2!}\mu_2 r^2 - \cdots} \tag{24}$$

where m_n is the nth moment of $\varphi(a)$ and the seminvariants μ can be computed from the moments by the algorithm

$$\begin{cases} m_1 = \mu_1 m_0 \\ m_2 = \mu_1 m_1 + \mu_2 m_0 \\ m_3 = \mu_1 m_2 + 2\mu_2 m_1 + \mu_3 m_0 \\ m_4 = \mu_1 m_3 + 3\mu_2 m_2 + 3\mu_3 m_1 + \mu_4 m_0 \end{cases} \tag{25}$$

etc.

In terms of these seminvariants the characteristic equation (18) becomes

$$\mu_1 r - \mu_2 \frac{r^2}{2!} + \cdots - \log_e m_0 = \log_e 1 = 2\pi n i \tag{26}$$

where n takes on all positive and negative integral values. Separating the real and imaginary parts in (26), and retaining seminvariants up to the fourth,

$$\psi(u, v) = \frac{\mu_4}{4!}(u^4 - 6u^2v^2 + v^4) - \frac{\mu_3}{3!}u(u^2 - 3v^2) + \frac{\mu_2}{2!}(u^2 - v^2) - \mu_1 u + \log_e m_0 = 0 \tag{27}$$

$$\chi(u, v) = \frac{\mu_4}{3!}uv(u^2 - v^2) + \frac{\mu_3}{3!}v(v^2 - 3u^2) + \mu_2 uv - \mu_1 v = 2\pi n. \tag{28}$$

If $\varphi(a)$ does not differ too widely from the normal (Gaussian) distribution, so that seminvariants of higher than second order can be neglected for roots in the neighborhood of $u = 0$, $v = 0$, we shall have, approximately[9]

$$\frac{\mu_2}{2!}(u^2 - v^2) - \mu_1 u + \log_e m_0 = 0 \tag{29}$$

$$\left(u - \frac{\mu_1}{\mu_2}\right)v = \frac{2\pi n}{\mu_2} \tag{30}$$

[9] The relations which follow hold *exactly* if $\varphi(a)$ is actually a normal curve. It should be noted, however, that this can not be strictly the case, since the infinite tail of the curve on the negative side would imply replacement or reproduction antedating the original installation or zero generation Nevertheless, a normal frequency curve will be admissible if the part of the curve extending into the negative age field is negligible. For a concrete example (electric light bulbs) see E. J. Gumbel, "Die Verteilung der Gestorbenen um das Normalalter," *Aktuarske Vedy* (Praze), 1933, p. 90.

or, putting

$$\left(u - \frac{\mu_1}{\mu_2}\right) = U \tag{31}$$

we have

$$U^2 - v^2 = \left(\frac{\mu_1}{\mu_2}\right)^2 - \frac{2 \log_e m_0}{\mu_2} \tag{32}$$

$$Uv = \frac{2\pi n}{\mu_2}. \tag{33}$$

It is thus seen that in these circumstances the roots u, v correspond to the points of intersection of the hyperbola (32) centered at $u = \frac{\mu_1}{\mu_2}$, $v = 0$, with a family of hyperbolas (33) concentric with (32), but with their axes at 45° to those of (32).

The intersections of the hyperbolas (33) with the axis of v are given by putting $u = 0$ in (30), namely

$$v = \frac{2\pi n}{\mu_1} \tag{30a}$$

This also gives, approximately, the frequency of the oscillatory components for which u is sufficiently small. In particular, for the first component, we have, in that case

$$v = \frac{2\pi}{\mu_1} \tag{30b}$$

so that its wave length is (approximately) μ_1, the mean of the $\varphi(a)$ curve.

These facts are illustrated in Fig. 1, drawn to scale according to the vital statistics of the United States, 1920, for which the requisite computations were available from prior publications (Lotka, (1928), (1929)). The diagram is drawn in full, showing four intersections of each hyperbola of the family (33). Actually values of v occur in pairs, corresponding to conjugate roots $u \pm iv$. The intersections in the two upper quadrants must be disregarded, as they do not correspond to roots of (18).

To simplify notation let us write (32), (33) in the form

$$U^2 - v^2 = K \tag{32a}$$

$$Uv = C. \tag{33a}$$

Solving for U^2, v^2 we find

$$U^2 = \tfrac{1}{2}\{K \pm \sqrt{K^2 + 4C^2}\} \tag{34}$$

$$v^2 = \tfrac{1}{2}\{-K \pm \sqrt{K^2 + 4C^2}\} \tag{35}$$

from which, incidentally, it is seen that

$$U^2 + v^2 = \sqrt{K^2 + 4C^2} \tag{36}$$

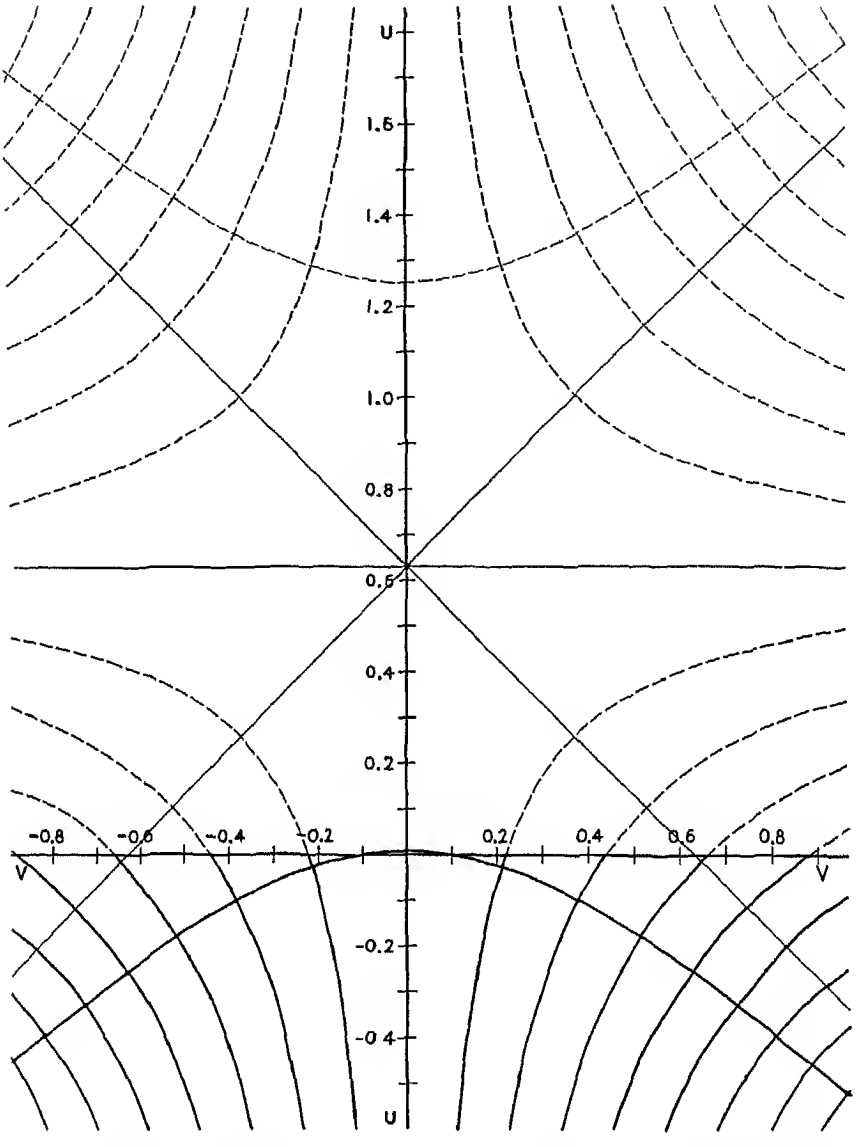


FIG. 1. ROOTS OF FUNDAMENTAL EQUATION (18) AS INTERSECTIONS OF CURVE (32) WITH FAMILY OF CURVES (33)

and hence, that the intersections of the hyperbola (32) with (33) lie on circles of radius

$$R = \sqrt[4]{K^2 + 4C^2}. \tag{37}$$

When the third and fourth moments (and therefore third and fourth seminvariants) are taken into account[10] the hyperbolas become distorted into new curves, though the general topographic features of the diagram tend to be preserved. In particular, the property of orthogonality of intersection of the curves (32) with (33) is preserved, in accordance with a well-known property of conjugate functions.[11] This is shown in the left hand panel of Fig 2, drawn for the same data as Fig. 1, but including not only the hyperbolic curves, but also the corresponding modified curves obtained by retaining the third and fourth seminvariants in the computation.[12] Only the quadrant relevant to the location of the roots is shown.

3. **The coefficients Q in the solution** (19). These are determined by initial conditions, being, in fact related to the function $B_1(t)$. As their determination in the original paper by Hertz and Herglotz is rather complicated, the following relatively simple method, resembling that by which the constants in a Fourier series are determined, is of interest:

Multiplying equation (9) by $e^{-r_i t}$, where r_i is a root of (18), transposing terms, and integrating between the limits 0 and ω, where ω is the highest age for which $\varphi(a)$ has a value other than zero, we have

$$\int_0^\omega e^{-r_i t} B_1(t)\, dt = \int_0^\omega e^{-r_i t}\left\{B(t) - \int_0^t B(t-a)\varphi(a)\, da\right\} dt. \tag{38}$$

Introducing the solution (19) in the right hand member of (38), we obtain

$$\int^\omega e^{-r_i t} B_1(t)\, dt = \sum Q_j \int_0^\omega e^{-r_i t}\left\{e^{r_j t} - \int_0^t e^{r_j(t-a)}\varphi(a)\, da\right\} dt \tag{39}$$

$$= \sum P_{ij} \qquad (j = 1, 2, 3, \cdots), \tag{40}$$

Consider now a particular term P_{ij} in the sum $\sum$. Multiplying out the exponentials we obtain

$$P_{ij} = Q_j \int_0^\omega e^{-(r_i - r_j)t}\left\{1 - \int_0^t e^{-r_j a}\varphi(a)\, da\right\} dt \tag{41}$$

which, in view of the characteristic equation (18) reduces to

$$P_{ij} = Q_j \int_0^\omega e^{-(r_i - r_j)t} \int_t^\omega e^{-r_j a}\varphi(a)\, da\, dt \tag{42}$$

$$= Q_j \int_0^\omega e^{-r_j a}\varphi(a) \int_0^a e^{-(r_i - r_j)t}\, dt\, da. \tag{43}$$

Hence, if $i \neq j$

$$P_{ij} = \frac{Q_j}{r_j - r_i} \int_0^\omega e^{-r_j a}\varphi(a)\{e^{-(r_i - r_j)a} - 1\}\, da \tag{44}$$

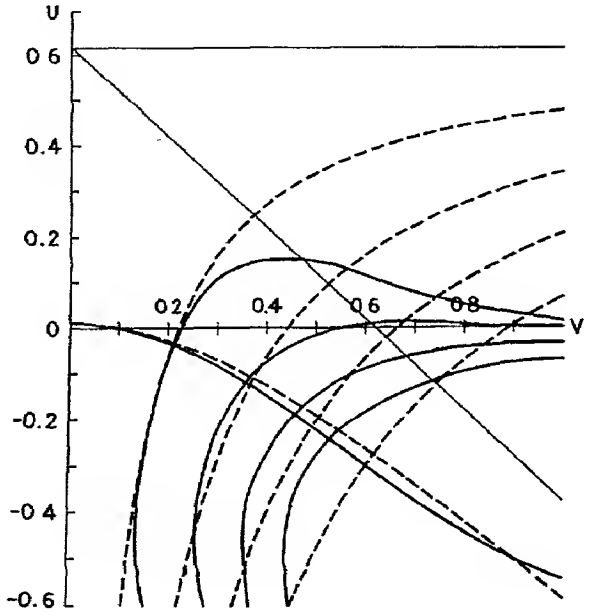
[10] Which is as far as curve fitting by Pearson's method goes.

[11] See, for example, W. E. Byerly, *Integral Calculus*, 1888, p. 289.

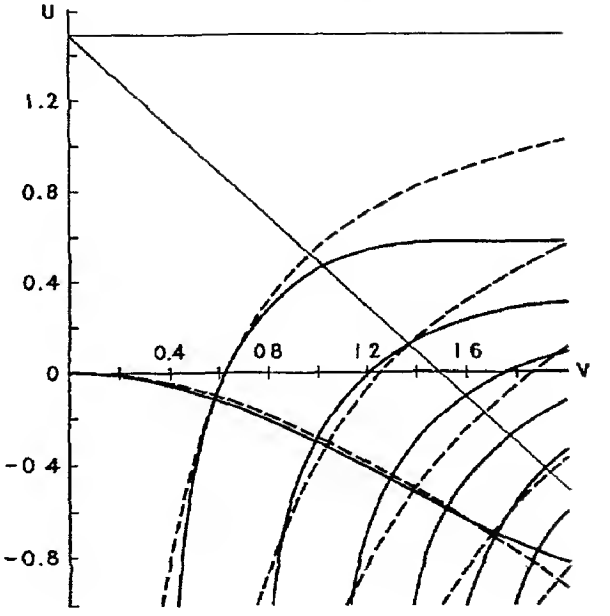
[12] For a given value of u equation (27) is a biquadratic in v, and equation (28) is a cubic in v lacking the second degree term The computation of the curves is in consequence relatively simple.

ROOTS OF FUNDAMENTAL EQUATION [18] AS INTERSECTIONS OF CURVE [27] OR [32] WITH FAMILY OF CURVES [28] OR [33] RESPECTIVELY

POPULATION GROWTH *



INDUSTRIAL REPLACEMENT †



——— Computed on basis of *first four* seminvariants, fourth degree equations [27], [28]
– – – – Computed on basis of *first two* seminvariants, equation of hyperbolas [32], [33]
† Data from Kurtz, E B, "Life Expectancy of Physical Property", 1930, page 104, fig 50.
* Data from Lotka, A J, "The Progeny of a Population Element", American Journal of Hygiene, 1928, page 875

$$= \frac{Q_j}{r_j - r_i}\left\{\int_0^\omega e^{-r_i a}\varphi(a)\,da - \int_0^\omega e^{-r_j a}\varphi(a)\,da\right\} \tag{45}$$

$$= 0 \tag{46}$$

since r_i and r_j are both roots of (18). But if $i = j$, then (44) is of the indeterminate form 0/0 and we must refer back to equation (43), from which, with $i = j$, we obtain, instead of (44) a different expression, namely

$$P_{ii} = Q_i \int_0^\omega e^{-r_i a}\varphi(a) \int_0^a dt\,da \tag{47}$$

$$= Q_i \int_0^\omega a e^{-r_i a}\varphi(a)\,da \tag{48}$$

so that the only term in the sum $\sum$ in equation (40) that does not vanish is the term P_{ii} and finally

$$Q_i = \frac{P_{ii}}{\int_0^\omega a e^{-r_i a}\varphi(a)\,da} \tag{49}$$

$$= \frac{\int_0^\omega e^{-r_i t}B_1(t)\,dt}{\int_0^\omega a e^{-r_i a}\varphi(a)\,da} \tag{50}$$

$$= \frac{\int_0^\omega e^{-r_i t}\left\{B(t) - \int_0^t B(t-a)\varphi(a)\,da\right\}dt}{\int_0^\omega a e^{-r_i a}\varphi(a)\,da} \tag{51}$$

or, finally, in view of (20)

$$Q_i = \frac{\int_0^\omega e^{-r_i t}\left\{B(t) - \int_0^t B(t-a)\varphi(a)\,da\right\}dt}{-\{F'(r)\}_{r=r_i}}. \tag{52}$$

The coefficients Q are thus fully determined by (50) or its equivalents (51) or (52), when initial conditions are given, that is, when the function $B_1(t)$ is given for $0 < t < \omega$ or, what amounts to the same thing, when $B(t)$ is known for this range of values of t. For complex roots the denominator in (52) becomes,[13] in view of (27), (28)

$$-\frac{dF(r)}{dr} = -\left\{\frac{\partial\psi}{\partial u} + i\frac{\partial\chi}{\partial u}\right\} = G - iH \tag{53}$$

[13] Since r_i is a root of $F(r) = 1$, we have

$$\left[\frac{dF(r)}{dr}\right]_{r=r_i} = \left[\frac{dF(r)}{F(r)\,dr}\right]_{r=r_i} = \left[\frac{d\log_e F(r)}{dr}\right]_{r=r_i}$$

where G and H can be expressed in terms of the seminvariants by partial differentiation of (27), (28) with regard to u, namely

$$G = \mu_1 - \mu_2 u + \frac{\mu_3}{2!}(u^2 - v^2) - \frac{\mu_4}{3!}(u^3 - 3uv^2) + \cdots \tag{54}$$

$$H = \mu_2 v - \mu_3 uv + \frac{\mu_4}{3!}(3u^2 v - v^3) - \cdots . \tag{55}$$

In the special case that the "zero generation" is composed of N_0 individuals (or "units") all born (or "entering") at time zero, the coefficients Q are correspondingly simplified in form. For the term in the real root r we have

$$Q = \frac{N_0}{-F'(r)} . \tag{56}$$

Conjugate complex root terms unite in pairs,[14] giving

$$Q'e^{(u+iv)t} + Q''e^{(u-iv)t} = \frac{2N_0 e^{ut}}{G^2 + H^2}\{G \cos vt - H \sin vt\}. \tag{57}$$

Unless $\varphi(a)$ is a normal distribution, the computation of the roots, u, v, and the coefficients G, H, in terms of seminvariants becomes impracticable for higher order roots, which then have to be computed directly and laborously from equations (22), (23). In practice components of very high order will hardly be needed, nor will their use be warranted, since the high order seminvariants, which are then involved, are not usually known with sufficient accuracy. An exception occurs when the $\varphi(a)$ curve is essentially of the nature of a composite curve. This is what actually happened in the case of the curve of reproduction for a human population. For details on this point the reader must be referred to my paper "The Progeny of a Population Element".

4. **Alternative Representation of the Function $B(t)$.** By the application of the Hertz-Herglotz solution of the integral equation (6), the evolution of a population or aggregate is represented as the resultant of a series of damped oscillations.

Additional insight into the nature of the renewal process is gained by viewing the total renewals as composed of contributions from successive "generations".[15]

---

[14] For details see A. J. Lotka, *The Progeny of a Population Element*, p. 892.

[15] In the case of a population the term "generation" calls for no explanation: mother, daughter and granddaughter, for example, represent three generations; in the case of industrial replacement, the term is to be understood in this sense, that the original installation constitutes the original or zero generation, the units introduced to replace disused units of the zero generation constitute the "first" generation, renewal of these the second, and so on.

This explanation may seem unnecessary. However, from some correspondence received by the writer it seems that perhaps some readers have confused the generations thus defined with successive "cycles" of duration equal to the extreme "length of life" of the units. With such "cycles" we are not here concerned.

This leads to an alternative representation, in which the evolution of the aggregate appears as the sum of a series of frequency curves, each corresponding to the contribution of one generation to the total births or replacements at time t.[16]

In order to realize this second representation we note, first of all, equation (7) applies not only to the total births at time t, but, with slight modification, also to the births in any particular generation. Here it will be convenient to consider the special case of a zero generation of N_0 individuals (or units) all born (or installed) at time $t = 0$.

The births (or renewals) in the "first" generation, that is offspring of the zero generation, or renewals of disused units of the zero generation, will be distributed in time according to the equation

$$B_1(t) = N_0\varphi(t). \tag{58}$$

For the second generation, or renewals of disused units of the first generation, we shall have

$$B_2(t) = \int_0^t B_1(t - a)\varphi(a)\, da \tag{59}$$

[16] This alternative approach of the problem bears some superficial resemblance to a method followed by R. Frisch in his article "Sammenhengen mellem primaerinvesteringen og reinvestering" (*Statsekonomisk Tidskrift*, 1927, p. 117) Frisch also follows up the distribution in time of first, second, and higher order replacements, and gives diagrams bearing a superficial resemblance to Fig. 4 in the present text. But Frisch's development has otherwise little in common with that here presented. He deals with equipment composed of various units, with expectation of life varying discontinuously or continuously from one unit to another, but fixed at a single value for a given unit. To use one of his own examples, it is as if a wooden hammer with a life of one year were always replaced by another wooden hammer, also with a life of one year, and so on: while a steel hammer, with a life of three years, were always replaced by another steel hammer, also with a life of exactly three years. The analogous case in population analysis would be presented by a population in which length of life were strictly hereditary, so that a man dying at age 50 would have a son, grandson, etc., each dying at age 50. In the field of industrial replacement and in population analysis alike this is a highly unrealistic supposition.

Needless to say, with these basic assumptions, Frisch's resulting equations differ fundamentally from those here given, and the distribution curves for successive orders of replacements, as shown in Frisch's Fig. 3 do *not* have the property that the j-th seminvariant of the k-th order replacement curve is k times that of the j-th seminvariant of the first order curve, except for $j=1$. The fact is that Frisch's curves in his Fig. 3 are all similar, except for a constant factor applied to the vertical scale and its reciprocal applied to the horizontal scale In this case all the corresponding seminvariants, except the first, are evidently unchanged in passing from one curve to the next. Frisch, as a matter of fact, does not introduce seminvariants into his discussion at all The Hertz solution he could not possibly introduce, since his fundamental equations are not of a form appropriate for the use of the Hertz solution.

The later sections of Frisch's paper deal with somewhat more complicated cases, but they all involve the assumption of "strict heredity," that is, the assumption that a unit with length of life v is replaced by another having exactly the same length of life v. At any rate, that is the understanding I have formed of the Danish text, studied with the assistance of a native of Scandinavia All the formulae in the text bear out this understanding.

and, generally, for the $(j + 1)$th generation[17]

$$B_{j+1}(t) = \int_0^t B_j(t - a)\varphi(a)\, da\,. \tag{60}$$

Now, by a well-known property[18] of the Thiele seminvariants, it follows from (58), (59), (60), that the seminvariants of the distribution-in-time of the births (or replacements) in the jth generation are simply the j-tuple of the corresponding seminvariants of the first generation, that is, of $\varphi(t)$.

Furthermore, it is easily shown that as j, the order of generation, increases, the distribution of renewals approaches[19] the normal (Gaussian) frequency distribution.

By virtue of these properties the distribution curves for successive generations are easily constructed.[20]

The sum total of the contributions of successive generations should, of course, agree with the expression for the total annual births $B(t)$ at time t given by the fundamental equation (9). In point of fact, by summing the left and the right hand members of equations (58), (59), and (60) for all generations up to the highest, say the n-th, "reproducing" at time t, we find

$$B(t) = \sum_{j=1}^{j=n+1} B_j(t) = B_1(t) + \int_0^t \sum_{j=1}^{j=n} B_j(t - a)\varphi(a)\, da. \tag{61}$$

Since the n-th is the highest generation contributing,[21] the value of the integral in (61) is not changed by writing n instead of $n + 1$ as the upper limit of the summation sign on the right. But then (61) becomes simply

$$B(t) = B_1(t) + \int_0^t B(t - a)\varphi(a)\, da$$

[17] The births in the j-th generation extend at most from $t = j\alpha$ to $t = j\omega$, but it is not necessary to take this into account in writing the limits of the integrals in (60) and corresponding equations, because the inclusion or exclusion of vanishing terms in the integrand does not affect the value of the integral. Similar remarks apply to the effect of the limited range of $\varphi(a)$. See also footnote 5.

[18] For details, see A. J. Lotka, "The Progeny of a Population Element," *American Journal of Hygiene*, 1928, vol. 8, p. 875; also "The Spread of Generations" *Human Biology*, 1929, vol 1, p. 305.

[19] In practice quite rapidly, even if $\varphi(a)$ is far from normal.

[20] For the case in which $\varphi(a)$ is a Pearson Type I curve, details of the process are given in my paper "Industrial Replacement," *Skandinavisk Aktuarietidskrift*, 1933, p. 51. I may here remark that such a Pearson Type I curve for the distribution in the first generation does not strictly give again a Pearson Type I curve in the second generation, because the moments beyond the 4th are neglected in fitting such a curve. But it must be remembered that the same neglect is practiced in the original fit of the data, so that the fit in the second generation will in general be as adequate as that in the first, provided, of course, that proper attention is paid to Pearson's criteria.

[21] The special case that the limiting n so defined is ∞ would require special discussion, which, however, presents no great difficulty. As this case is of little if any practical importance, this discussion is here omitted.

that is, summation of the contributions of individual generations to the total annual births, leads us back to the fundamental equation (9), which confirms the correctness of our analysis.

TABLE I

Age Schedule of Survivorship and of Replacements[22] *in First Generation*

Age Interval	Survivors from Original Installation to Beginning of Specified Age Interval	Replacements Within Specified Age Interval
0–1	100,000	—
1–2	100,000	—
2–3	100,000	300
3–4	99,700	900
4–5	98,800	1,800
5–6	97,000	3,000
6–7	94,000	5,700
7–8	88,300	10,300
8–9	78,000	14,100
9–10	63,900	13,900
10–11	50,000	13,800
11–12	36,200	13,200
12–13	23,000	10,400
13–14	12,600	6,300
14–15	6,300	3,700
15–16	2,600	2,200
16–17	400	400
17–18	—	—

5. **Application to Kurtz's data.** An extensive collection of numerical data (mortality curves) on renewal of industrial equipment has been published by E. B. Kurtz (1930), (1931). By way of example the analysis developed above has been applied to the data "Group III," as fitted by him with a Pearson Type I curve, namely[23]

$$B_1(t) = 14{,}950\left(1 + \frac{t-10}{12.67}\right)^{9.16}\left(1 - \frac{t-10}{10.43}\right)^{7.64}. \tag{62}$$

[22] Data from E. B. Kurtz, *Life Expectancy of Physical Property*, 1930, Table 22, Cols. 5 and 6, p. 86, and p. 104, Fig. 50.

[23] The numerical values of the constants in the formula as here given differ slightly from those given by Kurtz, perhaps owing to the retention by him of higher decimals in his computations. There is also an inconsistency between Kurtz's use of 10 for the mean in his formula, whereas on his drawing the mean is placed at 100.

The aperiodic component is the number of units originally installed (arbitrarily assumed as 100,000) divided by the mean of the frequency curve (equation 62). Following Kurtz, this has also been arbitrarily made equal to 10, which simply implies a particular choice of time unit. The fundamental data and characteristics are set forth in Tables I and II. The first six oscillatory components, were computed retaining moments and seminvariants up to μ_4, with the results shown in Table III and in Figs. 2 (right hand panel), 3 and 4.

TABLE II

Moments and Seminvariants of Curve of Replacements in First Generation[24]

j	Moments[25] m_j	Seminvariants μ_j
0	100,000	
1	0	10 [26]
2	671,924	6 7192
3	130,070	−1.3007
4	12,323,200	−12.1228

TABLE III

Constants of the Series Solution (19) of Integral Equation (7) for First Six Oscillatory Components Computed from First Four Moments and Seminvariants of an Industrial Replacement Curve[27]

Order of Component n	u	v	G	H	$\frac{G}{G^2+H^2}$	$\frac{H}{G^2+H^2}$
0	0	0	10.0000	0	.10000	0
1	−.11009	.57767	11 1688	4 1458	.07869	02921
2	−.30144	98920	14.3353	7 6696	05423	.02902
3	− 46500	1 28383	18 4982	10 4425	04100	.02314
4	− 59500	1.51475	23 1094	12 7773	.03314	01832
5	−.69800	1.70500	29 2088	14.8877	02718	01385
6	−.78000	1.86117	32 5165	16 7797	.02429	01253

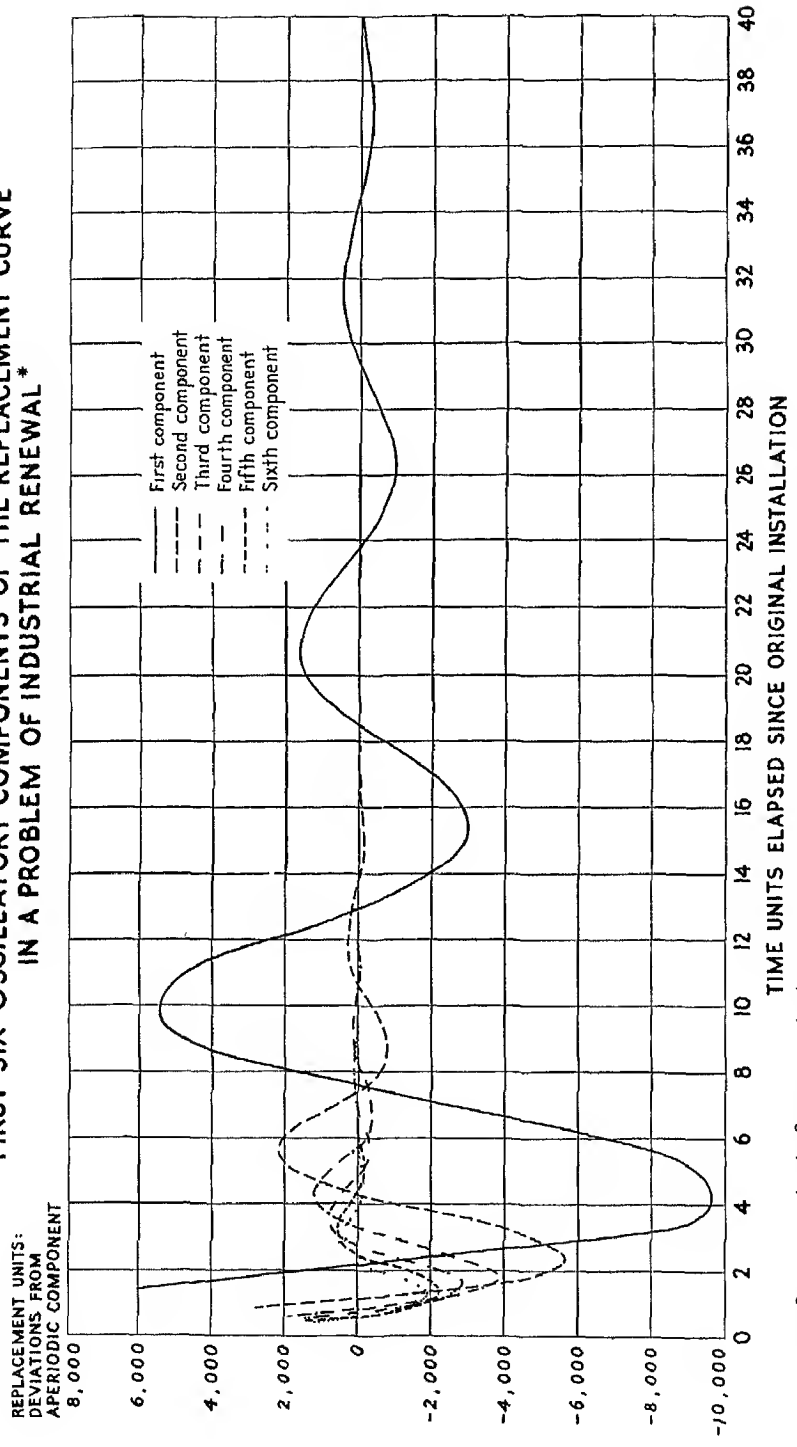
In particular, Fig. 4 shows the curve obtained by the summation of the first six oscillatory components superposed over the aperiodic (constant) component. It also shows the distribution curves of the first five generations within the range of the time scale on the diagram. Summation of these reproduces,

[24] Data from E. B. Kurtz, *Life Expectancy of Physical Property*, 1930, Table 22, p 86, and Fig. 50, p. 104.

[25] Moments taken about age 10

[26] This value of μ_1 is taken with reference to the origin.

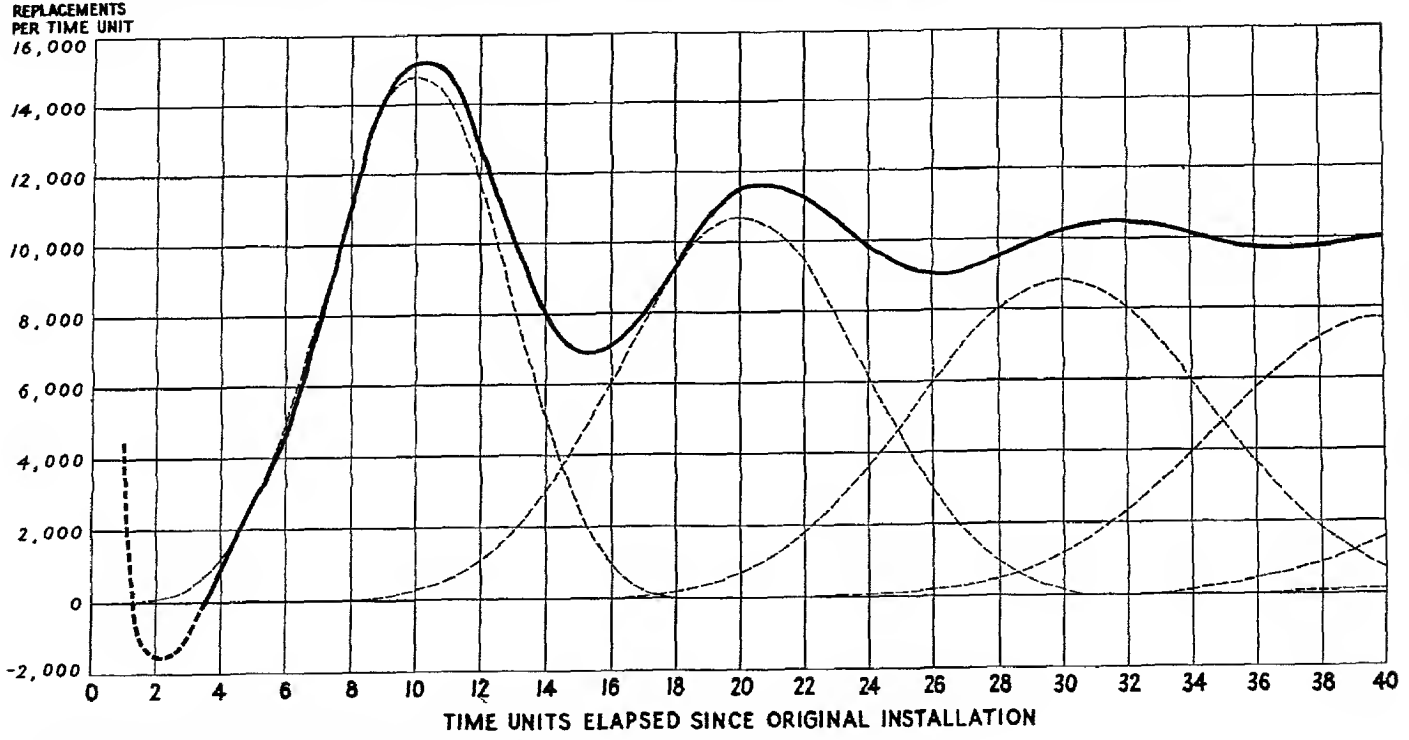
[27] Data from E. B Kurtz, *Life Expectancy of Physical Property*, 1930, p. 104, fig. 50



* Graph of solution (19) of equation (12) First 6 oscillatory components
Data from Kurtz, E B, "Life Expectancy of Physical Property", 1930, page 104, fig 50

FIG. 3

SUMMATION OF APERIODIC AND FIRST SIX OSCILLATORY COMPONENTS, AND FREQUENCY DISTRIBUTIONS OF SUCCESSIVE GENERATIONS OF REPLACEMENTS *



* Graph of Solution (19) of equation (12) summation of first six oscillatory components
Data from Kurtz, E B, "Life Expectancy of Physical Property", 1930, page 104, fig 50.

FIG. 4

within the errors of drawing, the resultant curve of the oscillatory solution, except for the very early stages of the process, where the oscillatory solution is of no practical interest, because the first generation alone dominates the whole process, and this is given by the observational data direct or after fitting with the curve such as (62).

It remains to consider briefly the relative advantages of the method of solution by differentiation, as originally applied by Herbelot, Risser, and others, on the one hand, and the use of the Hertz-Herglotz expansion, as introduced for the treatment of this type of problem by Sharpe and Lotka.

One obvious advantage of the method of differentiation *when it is applicable*, is that the result is obtained in the form of a closed, finite expression *for each cycle.*

Against this is to be reckoned, first, that the range of application of the method is severely limited. Preinreich in a recent issue of *Econometrica* (1938) uses for an illustration of the method a Pearson Type I curve, but in the very special and trivial form that the exponents are integers, namely 1 and 2. In practice the exponents will always be fractional, and then successive differentiations do not terminate as obligingly as in Preinreich's case. As already noted, Preinreich, though citing Kurtz's observational data on industrial replacement, discreetly abstains from using these for his numerical example.

Secondly, the disadvantage of a solution in form of an infinite series is more apparent than real. In practice the first few terms of the series obtained by the Hertz-Herglotz method will usually give an adequate representation of the facts, except for a short period immediately following the first installation. It is true that here this method, unless carried to high order components, may give an imperfect representation of industrial replacements, and may, in fact, give impossible negative values in this region, as in the example exhibited in Fig. 4. But this is practically unimportant, because in practice there will actually be few, if any, such very early replacements in an installation of finite dimensions. In fact, second and higher order replacements immediately after first installation are obviously out of the question in practice. For example, it may well happen once in a while that a telegraph pole is demolished on the very first day of service by collision with a truck. It is even imaginable that its replacement, put up the same day, might again be immediately demolished. But even in a country-wide installation one would hardly expect a third, fourth or fifth replacement to be required on the day of installation. In other words, that part of the replacement curve which relates to the very early period after first installation, is composed practically of first replacements only.

So for example in the diagram, Fig. 4, the curve of total replacements, up to about $t = 8$, is simply the curve of first replacements, which is given directly by the data of the problem. Within the range of errors of drawing the influence of higher components are quite unobservable in this region.

The case is even more favorable in the application of the method to the problem of population growth, for here there is actually no reproduction what-

ever until age α (say about 15) is reached. The part of the curve defined by the series (19) carried only to a finite number of terms,[28] and applied to values of $t < \alpha$, is therefore simply rejected.[29] It may save many words of explanation if the reader is simply referred to Fig. 4 on p. 897 of my previous publication "The Progeny of a Population Element," which illustrates the point, the minimum age of reproduction being just short of 15.

A major disadvantage of the method by differentiation is that it demands that the frequency distribution function $\varphi(a)$ be given in the form of a suitable analytic expression, or if it is not so given, that a *suitable* function or curve be fitted to it. The Hertz-Herglotz method, on the contrary, is directly applicable to the *raw data, regardless of their form.* Incidentally, curve fitting as practiced by Kurtz may produce a singular result. In 6 out of 7 of his types, the fitted frequency curve extends into negative field, implying that there are some replacements even before the actual installation. This may not be a very serious defect if the area of the curve in the negative field is negligible, but it should not pass unnoticed.

One of the principal merits of the Hertz-Herglotz expansion is that it renders the course of events over their whole extent, and, in particular, makes clear the mode of approach to the ultimate state represented by the aperiodic term. Because the method by differentiation requires a separate expression for each cycle, it is at best ill adapted to present to the eye or to the mind a comprehensive view of the evolution of the aggregate as a whole.

In the introductory paragraphs it was pointed out that the problems of population growth and those of industrial replacement were closely analogous, though there were certain points of difference. It is of interest here to give consideration to these differences.

One of these has already been noted. Replacement of industrial equipment may begin from the very moment of first installation, since accident as well as wear and tear must be provided for. Organic reproduction, on the other hand, does not occur immediately after birth. One result of this is that for any finite value of t, the number of generations contributing to the total births is itself finite; on the contrary, in the case of industrial replacement, if we interpret the equation (7) literally, there are at any moment an infinite number of generations contributing. In practice this, of course, does not occur, and the equation

[28] There are, of course, limitations to the application of the solution (19). No one with any experience in the treatment of practical problems by mathematical analysis would think of fitting, by means of a reasonably limited number of terms, the first phases of the processes here discussed, in the case of a rectangular distribution of the first generation, for example. But the distributions with which we are actually concerned in practice are far from rectangular. Such as they are, they are well adapted to the method, as is seen in the two examples illustrated.

[29] There is nothing unusual in this rejection of negative values of the frequency function where it falls outside the range of actual values. It is what we all do in using such a frequency curve as Pearson's type I, defined by a function which becomes negative outside the range of actual interest.

does not truly represent the facts in that a continuous distribution is assumed throughout, whereas for the higher order replacements ultimately the early frequencies are so thinned out that the discreteness of the units can no longer be disregarded.

Neverthcless, from the very start we must be prepared to consider several generations of replacement as contributing to the total; this lends a certain special interest, in dealing with the first cyclc of replacements, to the method of solution by differentiation, as used by Herbelot, Risser, Zwinggi, Schulthess, and lately Preinreich It is true that this interest is much diminished by the limitations in the applicability of the method.

On the other hand, in the case of organic reproduction, for the early part of the first cycle, the progeny of a population element belongs exclusively to a single ("first") generation. Between $t = 15$ and $t = 30$, in our example, only first generation births are taking place, and here the solution (19) is of more theoretical than practical interest, since the distribution of births is simply that of the first generation births.

Another point of difference is that the curve of $\varphi(a)$ in the case of industrial replacement, if we may judge by Kurtz's data, is a comparatively well behaved Pearson type curve. On the contrary, the corresponding curve of organic reproduction is a very inconvenient type to fit by any of the standard methods. In view of this it is all the more remarkable that the solution (19) gives as good a fit as it does with only four components, as will be seen on referring to my original publication, "The Progeny of a Population Element," p. 897, Fig. 4, already referred to.

Lastly, while the analogy is exact so long as we are dealing with industrial or organic aggregates maintained at a constant level, an essential difference arises when the case of a growing aggregate is considered. Organic growth takes place by what might be called "multiple replacement," that is, one individual in the course of life gives risc, on the average, to n individuals, where n may exceed unity. Analytically this finds expression in that

$$\int_0^\infty p(a)m(a)\,da > 1$$

and the fact is automatically taken care of in the solution (19) by the fact that in such a case the single real root $r > 0$.

Growth of industrial equipment, on the other hand, takes place by new units being installed *in addition to* replacement of disused units. The fundamental equations must be altered accordingly to take care of this case.

In conclusion I want to make a remark regarding the function of such analyses as the one here presented. In this connection I can do no better than to quote a sentence from Cournot:[30] "Those skilled in mathematical analysis know that

[30] A. Cournot, *Researches into the Mathematical Principles of the Theory of Wealth*, translated by N. Bacon, Macmillan Co., 1897, p. 3

its object is not simply to calculate numbers, but that it is also employed to find the relations between magnitudes . . . ."

It is essentially in this sense that the analysis of a problem of industrial replacement is here offered. If we are merely interested in numbers, the direct arithmetical approach as practiced by Kurtz may be as good as any. But if an insight into the anatomy of the processes involved, and into their evolution from an initial condition to a final state is desired, then the setting up of the fundamental equations, and their solution in exponential series or in other suitable analytical form, and a concise expression of the relation between the distributions in time of successive generations, or orders of replacements, have greatly superior merit as compared with brute attacks by arithmetic without regard to mathematical form. Nor are the systematic relations (in terms of certain seminvariants) that have been shown to exist between the distribution of successive generations to be regarded merely as "short cuts" for their computation, though sometimes they may be found convenient in that way. Their real significance lies in that they serve to complete for us the analytical picture of the process of evolution of the system under consideration.

METROPOLITAN LIFE INSURANCE COMPANY,
NEW YORK, N. Y.

BIBLIOGRAPHY

The following is a partial list of publications bearing on the general problem of self-renewing aggregates. The list is not the result of an exhaustive search, but merely of notes made in passing. Items marked with an asterisk are those referred to in the text. References are arranged essentially in chronological order.

1907 LOTKA, A J. "Relation between Birth Rates and Death Rates," *Science*, vol. 26, p 21.

"Studies of the Mode of Growth of Material Aggregates," *American Journal of Science*, vol 24, pp. 199, 375.

*1908 HERTZ, P. "Die Bewegung eines Elektrons unter dem Einflusse einer stets gleich gerichteten Kraft," *Mathematische Annalen*, vol. 65, p. 1.

HERGLOTZ, G. "Über die Integralgleichungen der Elektronentheorie," *Mathematische Annalen*, vol. 65, p. 87.

*1909 HERBELOT, L. "Application d'un Théorème d'analyse à l'étude de la repartition par âge," *Bulletin Trimestriel de l'Institut des Actuaires Français*, vol. 19; p. 293.

*1911 SHARPE, F. R. AND LOTKA, A. J. "A Problem in Age Distribution," *Philosophical Magazine*, vol. 21, p. 435.

*1912 RISSER, R. "Sur une application d'une équation fonctionnelle à un problème d'assurance," *Bulletin Trimestriel de l'Institut des Actuaires Français*, vol. 23, p. 298.

1913 LOTKA, A. J. "A Natural Population Norm," *Journal of the Washington Academy of Sciences*, vol. 3, pp. 241, 289.

1918 LOTKA, A. J. "The Relation between Birth Rate and Death Rate in a Normal Population and the Rational Basis of an Empirical Formula for the Mean Length of Life, given by William Farr," *Quarterly Publications of the American Statistical Association*, vol. 16, p. 121.

*1920 RISSER, R. "Sur une application de l'équation de Volterra au problème de la répartition par âge dans les milieux à effectif constant," *Comptes Rendus de l'Académie des Sciences*, p. 845.

BATEMAN, H. "An Integral Equation Occurring in a Mathematical Theory of Retail Trade," *Messenger of Mathematics*, vol. 49, p 134.

1921 MOSER, C. "Über Gleichungen für eine sich erneuernde Gesellschaft mit Anwendung auf Sozialversicherungskassen," *Verhandlungen der Schweizerischen Naturforschenden Gesellschaft*, p. 9.

LOTKA, A. J. "A Simple Graphic Construction for Farr's Relation between Birth Rate, Death Rate and Mean Length of Life," *Journal of the American Statistical Association*, vol. 17, p. 998.

1922 LOTKA, A. J. "The Stability of the Normal Age Distribution," *Proceedings of the National Academy of Sciences*, vol. 8, p. 339.

1923 DOETSCH, G. "Die Integrodifferentialgleichungen vom Faltungstypus," *Mathematische Annalen*, vol. 89, p. 192.

1924 SCHOENBAUM, E. "Anwendung der Volterra'schen Integralgleichungen in der mathematischen Statistik" I, II, III, *Skandinavisk Aktuarietidskrift*, p 241; 1925, p. 1.

FOCK, V. "Über eine Klasse von Integralgleichungen," *Mathematische Zeitschrift*, vol. 21, p. 161.

1925 LOTKA, A. J. *Elements of Physical Biology*, Williams & Wilkins Company, Baltimore, pp. 110, 117, 118 and 128.

DUBLIN, L. I. AND LOTKA, A. J. "On the True Rate of Natural Increase of a Population," *Journal of the American Statistical Association*, vol. 20, p. 569.

1926 LOTKA, A. J. "The Progressive Adjustment of Age Distribution to Fecundity," *Journal of the Washington Academy of Sciences*, vol. 16, p. 505.

NORTON, H. T. J. "Natural Selection and Mendelian Variation," *Proceedings of the London Mathematical Society*, Ser. 2, Part 1, vol. 28, p. 1.

MOSER, C "Beiträge zur Darstellung von Vorgängen und des Beharrungszustandes bei einer sich erneuernden Gesamtheit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 21, p. 1.

MEIDELL, B. Über periodische und angenäherte Beharrungszustände," *Skandinavisk Aktuarietidskrift*, No. 3–4, p. 172.

WICKSELL, S. D. "Sveriges framtida befolkning under olika förutsättningar," *Ekonomisk Tidskrift*, p. 91.

CONNOR, L. R. "Fertility of Marriage and Population Growth," *Journal of the Royal Statistical Society*, Part III, vol. 89, p. 553.

DOETSCH, G. "Bemerkung zu der Arbeit von V. Fock: Über eine Klasse von Integralgleichungen," *Mathematische Zeitschrift*, vol. 24, p. 785.

1927 LOTKA, A. J. "The Size of American Families in the Eighteenth Century," *Journal of the American Statistical Association*, vol. 22, p. 154.

"The Actuarial Treatment of Official Birth Records," *Eugenics Review*, vol. 19, p. 257.

TRICOMI, F. "Risoluzione di un problema demografico," *Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 22.

INSOLERA, F. "Sul calcolo della probabilità annuale di morte in gruppi aperti di popolazione, "*Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 611.

"Su particulari equazioni di Volterra e loro applicazione finanziaria e demografica," *Rendiconti del R. Instituto Lombardo di Scienze e Lettere*, p. 247.

FRIEDLI, W. "Über die Stabilität der gegenseitigen Hilfskassen," *Zeitschrift für Schweizerische Statistik und Volkswirtschaft*, No. 3, vol. 63, p. 322.

HALDANE, J. B. S. "A Mathematical Theory of Natural and Artificial Selection," *Cambridge Philosophical Society Proceedings*, vol. 23, p. 607.

FRISCH, R. "Sammenhengen mellem primaerinvestering og reinvestering," *Statsekonomisk Tidskrift*, p 117.

DOETSCH, G. "Überblick über Gegenstand und Methode der Funktionalanalysis," *Jahresbericht der deutschen Mathematiker-Vereinigung*, vol. 36, p. 1.

CANTELLI, F. P. "Intorno alla risoluzione di un problema demografico," *Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 449.

1928 MEIDELL, B. "Mathematical Statistics—On damping effects and approach to equilibrium in certain general phenomena," *Journal of the Washington Academy of Sciences*, No. 16, vol. 18, p. 437.

* LOTKA, A J "The Progeny of a Population Element," *The American Journal of Hygiene*, No. 6, vol. 8, p. 875.

1929 LOTKA, A. J. "The Spread of Generations," *Human Biology*, No. 3, vol. 1, p. 305.

"Biometric Functions in a Population Growing in Accordance with a Prescribed Law," *Proceedings of the National Academy of Sciences*, No. 10, vol. 15, p. 793.

WYSS, H. "Lage, Entwicklung und Beharrungszustand der eidgenossischen Versicherungskasse," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 24, p. 39.

1930 JENZER, H. "Hauptgrossen der Witwenversicherung bei Einfuhrung eines veränderlichen, exponentiellen Parameters fur Witwensterblichkeit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 25, p. 203.

WICKSELL, S. D. "Individens fruktsamhet och slaktets förokning," *Ekonomisk Tidskrift*, Häft 2, p 29.

KURTZ, E. B *Life Expectancy of Physical Property*, Ronald Press Company, New York

1931 ZWINGGI, E. "Die Witwenversicherung als Teil der allgemeinen Alters—und Hinterlassenenversicherung," *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, Heft 26, p 79.

"Das Problem der Erneuerung," *Festgabe Moser, Versicherungswissenschaftliche, juristische und mathematische Untersuchungen*, p 127.

HUSSON, R "Natalité et accroissement de la Population en France et à l'Étranger avant et après la guerre," *Bulletin de la Statistique générale de la France*, No 2, vol 20, p. 245.

WINFREY, R AND KURTZ, E B "Life Characteristics of Physical Property," *Iowa State College, Bulletin* 103, vol 30, No. 3.

ZWINGGI, E. "Zum Problem der Erneuerung," *Blatter fur Versicherungs-Mathematik und verwandte Gebiete*, No. 1, vol 31, p 18.

WYSS, H "Über die Pramiensysteme der Gruppenversicherung," *Festgabe Moser*, p. 175.

WICKSELL, S. D. "Nuptiality, Fertility, and Reproductivity," *Skandinavisk Aktuarietidskrift*, p. 125

BONZ, F. AND HILBURG, F. "Die voraussichtliche Bevölkerungsentwicklung in Deutschland," *Zeitschrift fur angewandte Mathematik und Mechanik*, No. 3, vol. 11, p 237.

GOLDSTEIN, S. "Note on the Operational Solution of an Integral Equation," *Journal of the London Mathematical Society*, Part 4, vol 6, p. 262.

LOTKA, A. J. "The Structure of a Growing Population," Paper presented before the *Second General Assembly of the International Union for the Scientific Investigation of Population Problems*, London, 1931 and published in the *Proceedings* (Problems of Population) George Allen & Unwin, 1932. Also published in *Human Biology*, vol. 3, p. 459.

*1933 LOTKA, A. J. "Industrial Replacement," *Skandinavisk Aktuarietidskrift*, 1933, p. 51.

"Applications de l'analyse au phénomène démographique," *Journal de la Société de Statistique de Paris*, vol. 74, p. 336.

ZWINGGI, E "Methodischer Beitrag zur Deckungs-Kapitalberechnung in der Sozialversicherung," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 28, p. 77.

v. MISES, R. Über die Vorausberechnung von Umfang und Altersschichtung der Bevölkerung Deutschlands," *Blätter für Versicherungs-Mathematik und verwandte Gebiete*, No. 10, vol. 2, p. 359.

1934 MARET, A. *Untersuchungen über diskontinuierlich sich erneuernde Gesamtheiten.* Inaugural Dissertation, Bern.

WICKSELL, S. D "Bidrag till den formella befolkningsteorien," *Statsokonomisk Tidsskrift*, Heft 1–2, p. 1.

*1935 SCHULTHESS, H. *Volterra'sche Integralgleichungen in der Versicherungsmathematik.* Inaugural Dissertation der Philosophischen Fakultat II der Universitat Bern.

1936 LOTKA, A J. "The Geographic Distribution of Intrinsic Natural Increase in the United States, and an Examination of the Relation between Several Measures of Net Reproductivity." *Journal of the American Statistical Association*, vol. 31, p. 273.

LOTKA, A. J. "Modern Trends in Birth Rate," *The Annals of the American Academy of Political and Social Science*, vol 188, November, p. 1.

SCHULTHESS, H. "Über die Verwendung von Integralgleichungen zur Berechnung der mathematischen Reserve (Deckungskapital)," *Blätter für Versicherungs Mathematik und verwandte Gebiete*, vol. 3, p. 412.

KERMACK, W. O. AND MCKENDRIK, A. G. "The Solution of Sets of Simultaneous Integral Equations Related to the Equation of Volterra," *Proceedings of the London Mathematical Society*, Ser. 2, vol. 41, p. 462; 472–477.

KREIS, H. "Stabilitat einer sich jährlich erneuernden Gesamtheit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 32, p. 17.

1937 LOTKA, A. J. "A Historical Error Corrected," *Human Biology*, vol. 9, No. 1, p. 104.

"Population Analysis: A Theorem Regarding the Stable Age Distribution," *Journal of the Washington Academy of Sciences*, vol. 27, p. 299.

SCHULTHESS, H. "Über das Erneuerungsproblem bei Verwendung eines analytischen Sterbegesetzes," *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, Heft 33, p. 69

HADWIGER, H. "Zur Berechnung der Erneuerungsfunktion nach einer Formel von V. A Kostitzin," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 34, p. 37.

LOTKA, A. J. "Quelques résultats récents de l'analyse démographique," *Congrès International de la population*, Paris. I. Théorie générale de la population, p. 96.

HUSSON, R "La démographie mathématique, Équations générales Propriétés générales des trajectoires démographiques," *Congrès International de la Population*, Paris. I. Théorie générale de la population, p. 108.

1938 LOTKA, A. J. "Some Recent Results in Population Analysis," *Journal of the American Statistical Association*, vol. 33, p. 164.

*1938 PREINREICH, G A. D. "Annual Survey of Economic Theory: The Theory of Depreciation," *Econometrica*, vol. 6, p 219

ZWINGGI, E. "Bemerkungen zum Erneuerungsproblem," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 36, p 69.

HADWIGER, H. "Über die Entwicklung der Erneuerungsfunktion in eine Potenzreihe," *Skandinavisk Aktuarietidskrift*, Haft 1–2, p. 31.

ON THE MATHEMATICS OF THE REPRESENTATIVE METHOD OF SAMPLING[1]

By Allen T. Craig

1. **Introduction.** This paper is designed to present certain topics in mathematical statistics which find application in some of the problems that arise in what has been termed the representative method of sampling.

For descriptive purposes, it seems convenient to consider two aspects of the representative method. The first of these may be called the method of *purposive selection.* This method can be roughly characterized by saying that it is the method employed when the samples are chosen in such a way that each sample will possess one or more characters, say certain averages, which are identical with the corresponding characters in the population from which the samples are drawn. The mathematical conditions which underlie this method are rather stringent, and both theoretical and practical investigations seem to have proved that in general no great amount of confidence can be placed in the results obtained.

The second aspect of the representative method has been styled the method of *random sampling.* This method can take either of two forms which we may call the method of *unrestricted random sampling* and *stratified random sampling.* The first of these is the classical method of procedure. That is, a sample is drawn at random from a given population and on the basis of these data inferences are made concerning the nature of the population. On the other hand, when the method of stratified random sampling is used, the population is first separated into a large number of parts, called strata, and the sample consists of an equally large number of "partial samples," each partial sample being drawn from a different stratum. It appears, both from theoretical and practical results, that this method of stratified random sampling enjoys many advantages not shared by the other methods.

We now turn to the main purpose of this paper, namely that of enumerating some of the theorems and methods of mathematical statistics which serve useful purposes in this theory. Discussion of how these theorems find application in the method itself has been reserved for other participants on this program.

2. **Estimates.** From our preliminary remarks, it is apparent that the representative method is much concerned with the problem of estimating certain

[1] Presented, at the invitation of the program committee, to a joint session of the Institute of Mathematical Statistics and the American Statistical Association on December 29, 1938.

unknown *parameters* of a statistical population. On this account, we first consider the problem of estimates.

Consider a population with arithmetic mean m and standard deviation σ. Let $x_1, x_2, \cdots, x_n$, be n independent items drawn from this population and let $c_1, c_2, \cdots, c_n$ be any finite real constants, not all zero to avoid the trivial case. Write $y = c_1x_1 + c_2x_2 + \cdots + c_nx_n$. Then the expected or arithmetic mean value of y is

$$\bar{y} = E(y) = m(c_1 + c_2 + \cdots + c_n),$$

and the variance of y is

$$\sigma_y^2 = E\{(y - \bar{y})^2\} = \sigma^2(c_1^2 + \cdots + c_n^2).$$

Suppose we inquire into the probability that y will have a value which is within a preassigned ϵ of its expected value. To this end, let C be the numerical value of the numerically greatest of the set $c_1, \cdots, c_n$, so that $\sigma_y^2 \leq n\sigma^2C^2$. Then by Tchebycheff's inequality p, the probability that $| y - \bar{y} | < \epsilon$, where ϵ is an arbitrarily small positive number, is such that

$$p \geq 1 - \frac{\sigma_y^2}{\epsilon^2},$$

or

$$p \geq 1 - \frac{n\sigma^2C^2}{\epsilon^2}.$$

In general, this inequality will have little interest. But if C is of the form $M/n^{\frac{1+\delta}{2}}$, M independent of n, $\delta > 0$, then $p \geq 1 - \frac{\sigma^2M^2}{n^\delta\epsilon^2}$ and by increasing n the right member can be made as near to one as we please. This means then that if we have a population with a finite variance and if we construct a linear function of the observations with coefficients of the nature indicated, we can, by increasing the size of the sample, make the probability approach one that the linear function will have a value arbitrarily close to its expected value.[2]

Now suppose that instead of constructing an arbitrary linear function we attempt to construct a function which will be an estimate of some particular parameter of the population. If the estimate is to be most serviceable, we should like to be able, by governing the size of the sample, to be as certain as we like that the estimate will have a value arbitrarily near that of the parameter. The preceding discussion shows that we can best achieve this by requiring that the expected value of the estimate be equal to the parameter sought. An estimate such as that just described is frequently called an *unbiased* estimate. The use of such estimates in statistical problems makes it possible to avoid systematic errors in estimating parameters. In general, unique unbiased estimates of a parameter do not exist. For example, the arithmetic mean m of

[2] Under these conditions, the function of the observations is said to converge stochastically to its expected value.

the population can be estimated from the sample $x_1, \cdots, x_n$ by any one of a large number of unbiased estimates such as $(x_1 + x_2 + \cdots + x_n)/n$, $(x_1 + x_n)/2$, x_4, and so on without limit. Thus it becomes necessary to make a choice of the unbiased estimate to be used. An appropriate criterion is that the unbiased estimate whose distribution has the smallest variance is the best to use. The reason for this can be seen by examining the preceding formula $p \geq 1 - \frac{\sigma_y^2}{\epsilon^2}$. For if y_1 and y_2 are two unbiased estimates of the same parameter and if $\sigma_{y_1}^2 < \sigma_{y_2}^2$, then in $p_1 \geq 1 - \frac{\sigma_{y_1}^2}{\epsilon^2}$ and $p_2 \geq 1 - \frac{\sigma_{y_2}^2}{\epsilon^2}$ we see that $1 - \frac{\sigma_{y_1}^2}{\epsilon^2}$ is more nearly equal to one than is $1 - \frac{\sigma_{y_2}^2}{\epsilon^2}$. Because of this fact we prefer, at least in most problems, to use y_1 rather than y_2 as an estimate of the unknown parameter. An unbiased estimate whose sampling variance is a minimum is sometimes called a *best estimate.*[3] It should not be inferred that the word "best" has any implications other than those stated explicitly in the definition.

The question very naturally arises as to whether we can determine these best estimates in particular cases. In general we can not determine them, but under certain conditions we can find best estimates if we are dealing with linear functions of the observations. A method and the conditions are set forth in an important theorem due to Markoff. We now consider his method.

3. Markoff's Method. Let there be given n statistical populations with arithmetic means $m_1, m_2, \cdots, m_n$ and standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_n$ respectively. We assume that no correlation exists between any of the populations. Furthermore, suppose that each of the n arithmetic means can be expressed linearly in terms of k unknown, but unique, parameters, say $z_1, z_2, \cdots, z_k$. Thus

$$
\begin{aligned}
m_1 &= a_{11}z_1 + a_{12}z_2 + \cdots + a_{1k}z_k \\
m_2 &= a_{21}z_1 + a_{22}z_2 + \cdots + a_{2k}z_k \\
&\vdots \\
m_n &= a_{n1}z_1 + a_{n2}z_2 + \cdots + a_{nk}z_k,
\end{aligned} \tag{1}
$$

where the a's are known constants. Likewise, let T be a parameter which is expressible linearly in terms of the same k unknown parameters, say $T = b_1z_1 + b_2z_2 + \cdots + b_kz_k$, where the b's are given constants. We draw a sample of n independent items, $x_1, x_2, \cdots, x_n$, in which one item is drawn from each

[3] An estimate of a parameter which converges stochastically (cf. footnote (2)) to that parameter is called a *consistent* estimate of the parameter If a consistent estimate has a distribution which is normal for large samples and if the variance of that distribution is smaller than the variance of any other consistent estimate which also has a normal distribution for large samples, then the estimate is called *efficient.* It should be observed that our definition of best estimate requires an unbiased estimate, whereas consistent and efficient estimates may be biased.

of the n populations. On the basis of this sample we seek to determine a set of numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$ such that $T' = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n$ is the best estimate of T.

Before attempting to find the solution, if one exists, let us first examine the mathematical implications of the problem. In the first place, in order that parameters $z_1, \cdots, z_k$ may exist, it is necessary and sufficient that the matrices A and B, where

$$A = \begin{Vmatrix} a_{11} & a_{12} \cdots a_{1k} \\ a_{21} & a_{22} \cdots a_{2k} \\ \vdots & \\ a_{n1} & a_{n2} \cdots a_{nk} \end{Vmatrix} \text{ and } B = \begin{Vmatrix} a_{11} & a_{12} \cdots a_{1k} & m_1 \\ a_{21} & a_{22} \cdots a_{2k} & m_2 \\ \vdots & \\ a_{n1} & a_{n2} \cdots a_{nk} & m_n \end{Vmatrix}$$

have the same rank. Thus we require that A and B have the common rank R. This being satisfied, we note further that if $k > n$, there will be infinitely many values of the z's which will satisfy the equations (1). Thus we require in addition that $k \leq n$. Finally, we note that if the common rank R is less than k, there will be infinitely many values of the z's which will satisfy the system (1). Hence we must have $R = k \leq n$.

We now turn to a consideration of the solution of the problem. Whatever the values of the λ's, we have for the mean value and the variance of T'

$$\begin{aligned} E(T') &= \lambda_1 m_1 + \cdots + \lambda_n m_n \\ &= \lambda_1 \Sigma a_{1j} z_j + \cdots + \lambda_n \Sigma a_{nj} z_j , \end{aligned}$$

and

$$\sigma_{T'}^2 = \lambda_1^2 \sigma_1^2 + \cdots + \lambda_n^2 \sigma_n^2 ,$$

respectively. Since $E(T')$ must equal T as a part of the condition for a best estimate, then

$$\lambda_1 \Sigma a_{1j} z_j + \cdots + \lambda_n \Sigma a_{nj} z_j = b_1 z_1 + \cdots + b_k z_k$$

identically in the z's. That is, the coefficients of $z_1, \cdots, z_k$ in the left member must equal the corresponding coefficient in the right member. Accordingly,

$$\begin{aligned} a_{11}\lambda_1 + a_{21}\lambda_2 + \cdots + a_{n1}\lambda_n &= b_1 \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \cdots + a_{n2}\lambda_n &= b_2 \\ &\vdots \\ a_{1k}\lambda_1 + a_{2k}\lambda_1 + \cdots + a_{nk}\lambda_n &= b_k . \end{aligned} \tag{2}$$

If these equations are to have solutions for $\lambda_1, \cdots, \lambda_n$, we must make the additional assumption that the matrix C, where

$$C = \begin{Vmatrix} a_{11} & a_{21} \cdots a_{n1} & b_1 \\ a_{12} & a_{22} \cdots a_{n2} & b_2 \\ \vdots & \\ a_{1k} & a_{2k} \cdots a_{nk} & b_k \end{Vmatrix}$$

has the same rank as the matrix of the coefficients, namely R. If this condition is satisfied we can write equations (2) in the form

$$\begin{aligned} a_{11}\lambda_1 + \cdots + a_{k1}\lambda_k &= b_1 - a_{k+1,1}\lambda_{k+1} - \cdots - a_{n1}\lambda_n \\ &\vdots \\ a_{1k}\lambda_1 + \cdots + a_{kk}\lambda_k &= b_k - a_{k+1,k}\lambda_{k+1} - \cdots - a_{nk}\lambda_n \end{aligned} \tag{3}$$

and solve for $\lambda_1, \cdots, \lambda_k$ in terms of the a's, the b's, and $\lambda_{k+1}, \cdots, \lambda_n$. Here, without any essential loss of generality, we take the non-vanishing k-rowed determinant to be that of the coefficients of $\lambda_1, \cdots, \lambda_k$ in equations (2). Thus for arbitrarily assigned values of $\lambda_{k+1}, \cdots, \lambda_n$, we can compute the values of $\lambda_1, \cdots, \lambda_k$ and these n values of the λ's will give us a T' which is an unbiased estimate of T. That there will be, in general, an unlimited number of sets of values of the λ's is in keeping with our previous observation that unique unbiased estimates usually do not exist.

The next part of the problem will consist in determining which, if any, of the above sets of λ's will make $\sigma^2_{T'}$ a minimum. We recall that $\sigma^2_{T'} = \lambda_1^2\sigma_1^2 + \cdots + \lambda_n^2\sigma_n^2$. In $\sigma^2_{T'}$ let us replace $\lambda_1, \cdots, \lambda_k$ by their values (in terms of $\lambda_{k+1}, \cdots, \lambda_n$) which we obtained by solving the system (3). Then $\sigma^2_{T'}$ will be expressed in terms of $\sigma_1, \cdots, \sigma_n$, the a's, the b's, and $\lambda_{k+1}, \cdots, \lambda_n$. We next take the partial derivative of $\sigma^2_{T'}$ with respect to each of $\lambda_{k+1}, \cdots, \lambda_n$. On equating these partial derivatives to zero we will have a system of $n - k$ linear equations in the $n - k$ unknowns $\lambda_{k+1}, \cdots, \lambda_n$. If these equations yield unique values for $\lambda_{k+1}, \cdots, \lambda_n$, they will in turn determine unique values of $\lambda_1, \cdots, \lambda_k$. This gives us a unique set of λ's such that at one and the same time

$$E(T') = T \text{ and } \sigma^2_{T'} \text{ is a minimum.}$$

The procedure which we have just outlined is most tedious to carry out in a particular case. Because of the insight of Markoff, a much better scheme is available for finding the best estimate of T. Consider the function of $z_1, \cdots, z_k$,

$$\begin{aligned} F(z_1, \cdots, z_k) &= \sum \left(\frac{x_j - m_j}{\sigma_j}\right)^2 \\ &= \sum \left(\frac{x_j - a_{j1}z_1 - \cdots - a_{jk}z_k}{\sigma_j}\right)^2. \end{aligned}$$

Evaluate $\dfrac{\partial F}{\partial z_1}, \cdots, \dfrac{\partial F}{\partial z_k}$ and equate these partial derivatives to zero. This yields the following system of k linear equations in the k unknowns $z_1, \cdots, z_k$.

$$\begin{aligned} z_1 \sum \frac{a_{j1}^2}{\sigma_j^2} + \cdots + z_k \sum \frac{a_{j1}a_{jk}}{\sigma_j^2} &= \sum \frac{a_{j1}x_j}{\sigma_j^2} \\ &\vdots \\ z_1 \sum \frac{a_{j1}a_{jk}}{\sigma_j^2} + \cdots + z_k \sum \frac{a_{jk}^2}{\sigma_j^2} &= \sum \frac{a_{jk}x_j}{\sigma_j^2}. \end{aligned} \tag{4}$$

If the system (4) yields unique values for the z's, these values, when substituted in T, yield exactly the same estimate of T as was found by substituting for the λ's in T'.

Perhaps an illustration will make this clearer. Suppose we have $n = 2$ populations and that the means m_1 and m_2 are expressible linearly in terms of $k = 1$ parameter z_1. Our equations (1) become

$$\begin{aligned} m_1 &= a_{11}z_1 \\ m_2 &= a_{21}z_1 . \end{aligned} \tag{1'}$$

Similarly, we have $T = b_1z_1$ and $T' = \lambda_1x_1 + \lambda_2x_2$. We first determine the λ's such that T' is the best estimate of T. In accordance with the preceding steps, equations (2) become

$$a_{11}\lambda_1 + a_{21}\lambda_2 = b_1, \tag{2'}$$

and the system (3) becomes

$$\lambda_1 = \frac{b_1 - a_{21}\lambda_2}{a_{11}}, \qquad a_{11} \neq 0. \tag{3'}$$

Then

$$\begin{aligned} \sigma^2_{T'} &= \lambda_1^2\sigma_1^2 + \lambda_2^2\sigma_2^2 \\ &= \left(\frac{b_1 - a_{21}\lambda_2}{a_{11}}\right)^2 \sigma_1^2 + \lambda_2\sigma_2^2, \end{aligned}$$

because of (3'). Thus

$$\frac{\partial\sigma^2_{T'}}{\partial\lambda_2} = \frac{-2a_{21}(b_1 - a_{21}\lambda_2)\sigma_1^2}{a_{11}^2} + 2\lambda_2\sigma_2^2,$$

and for a minimum $\sigma^2_{T'}$ we write $\frac{\partial\sigma^2_{T'}}{\partial\lambda_2} = 0$ so that

$$\lambda_2 = \frac{a_{21}b_1\sigma_1^2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

Since

$$\lambda_1 = (b_1 - a_{21}\lambda_2)/a_{11},$$

then

$$\lambda_1 = \frac{b_1a_{11}\sigma_2^2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

Our best estimate of T is found from T' and it is

$$T' = \frac{b_1a_{11}\sigma_2^2x_1 + b_1a_{21}\sigma_1^2x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

By Markoff's method we would form the function

$$F(z_1) = \sum \left(\frac{x_j - a_{j1}z_1}{\sigma_j}\right)^2 = \left(\frac{x_1 - a_{11}z_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - a_{21}z_1}{\sigma_2}\right)^2.$$

The system (4) reduces to merely

$$z_1 = \frac{a_{11}\sigma_2^2 x_1 + a_{21}\sigma_1^2 x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}. \tag{4'}$$

We substitute this value of z_1 in $T = b_1 z_1$ and obtain

$$T = \frac{b_1 a_{11}\sigma_2^2 x_1 + b_1 a_{21}\sigma_1^2 x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2},$$

which is the estimated value T' above.

4. **Neyman's modification of Markoff's Method.** We are indebted to Neyman for a modification and adaptation of the Markoff method so as to make the method applicable to some of the problems of stratified random sampling. One of his examples will best illustrate the method.

Suppose that a given population is divided into n strata. Let the jth stratum contain M_j items and let these items be $u_{j1}, u_{j2}, \cdots, u_{jM_j}$. The mean and the variance of this stratum are then

$$\bar{u}_j = \frac{1}{M_j}\sum_k u_{jk} \quad \text{and} \quad \sigma_j^2 = \frac{1}{M_j}\sum_k (u_{jk} - \bar{u}_j)^2.$$

Let T be the parameter $T = M_1\bar{u}_1 + M_2\bar{u}_2 + \cdots + M_n\bar{u}_n$, so that $\frac{T}{M_1 + \cdots + M_n}$, the mean of the population, is expressed as a linear function of the means of the n strata. We draw at random a sample of N items, the sample consisting of n partial samples, one partial sample being drawn from each of the n strata. Suppose there are n_1 items in the partial sample from the first stratum, n_2 from the second, and so on. Thus $n_1 + n_2 + \cdots + n_n = N$ and the entire sample consists of the n partial samples

$$x_{11}, x_{12}, \cdots, x_{1n_1}$$
$$x_{21}, x_{22}, \cdots, x_{2n_2}$$
$$x_{n1}, x_{n2}, \cdots, x_{nn_n}.$$

From these N data we propose constructing an estimate

$$T' = \lambda_{11}x_{11} + \cdots + \lambda_{1n_1}x_{1n_1} + \cdots + \lambda_{n1}x_{n1} + \cdots + \lambda_{nn_n}x_{nn_n}$$

which will be the best estimate of T. Now the expected value of T' is

$$\begin{aligned} E[T'] = E\Big[\sum_j \sum_k \lambda_{jk}x_{jk}\Big] &= \sum_j \sum_k \lambda_{jk}E(x_{jk}) \\ &= \sum_j \sum_k \lambda_{jk}\bar{u}_j \\ &= \sum_j \bar{u}_j \sum_k \lambda_{jk}, \end{aligned}$$

which, by hypothesis, must equal T. Thus

$$\sum_1^n \bar{u}_j \sum_1^{n_j} \lambda_{jk} = \sum_1^n M_j \bar{u}_j$$

identically in the $\bar{u}$'s. Hence $\Sigma \bar{u}_j(M_j - \Sigma\lambda_{jk}) = 0$ which requires that the coefficients of $\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_n$ must be zero. That is

$$\sum_1^{n_1} \lambda_{1k} = M_1$$

$$\vdots$$

$$\sum_1^{n_n} \lambda_{nk} = M_n.$$

Of course there are infinitely many λ's which will satisfy these equations. But we can eliminate all but one set by imposing the condition that $\sigma^2_{T'}$ shall be a minimum. The algebra of mathematical expectation can be used to show that

$$\sigma^2_{T'} = \sum_1^n \sigma_j^2 \left[\frac{M_j n_j - n_j^2}{M_j - 1} \left(\frac{1}{n_j} \sum \lambda_{jk} \right)^2 + \frac{M_j}{M_j - 1} \sum \left(\lambda_{jk} - \frac{1}{n_j} \sum \lambda_{jk} \right)^2 \right]$$

which will be a minimum when $\sum \left(\lambda_{jk} - \frac{1}{n_j} \sum \lambda_{jk} \right)^2 = 0, j = 1, 2, \cdots, n$. Since this is a sum of real squares, each term in the sum must be zero. Thus, $\lambda_{jk} = \frac{1}{n_j} \sum \lambda_{jk}$. Since $\sum \lambda_{jk}$ must equal M_j in order that $E(T') = T$, then $\lambda_{jk} = \frac{M_j}{n_j}$ which uniquely determines the λ's and hence our best estimate of T'.

It is important to observe that Neyman's adaptation does not assume that the various strata are uncorrelated nor that there are necessarily replacements after each drawing in taking the sample.

5. **Estimation of Ratios.** In certain problems in representative sampling it may be necessary to estimate both the numerator and the denominator of a fraction, say T/U. If T' and U' are linear estimates of T and U then for large samples both T' and U' will be approximately normally distributed in most cases. Further, if T' and U' are correlated, they will usually be approximately normally correlated. Geary has proved that if we write

$$V = \frac{b + T'}{a + U'},$$

where a and b are constants and U' and T'' are measured from their expected values, then

$$t = \frac{aV - b}{\sqrt{V^2\sigma^2_{U'} - 2rV\sigma_{T'}\sigma_{U'} + \sigma^2_{T'}}}$$

is approximately normally distributed with mean zero and unit variance provided $a \geq 3\sigma_{U'}$. Here r is the correlation coefficient between T' and U'. For

large samples this provides a convenient method of testing the significance of the difference between an observed and a hypothetical ratio of two linear estimates.

6. **Fiducial Inference.** After an estimate of a parameter has been made, it is usually desirable to make some inference about the true value of the parameter. For many years the concept of probable error was used in this connection. But the use of the probable error involves the assumption that all values of the unknown parameter are equally likely. This assumption is questionable and efforts to avoid making the assumption have led to a theory called *fiducial inference*. This method of statistical inference has broad implications but limitations on our time do not permit our discussing the topic. At the close of this paper, we give certain references to the subject, including some of an expository nature.

7. **Conclusion.** As stated in the introduction, this paper purports to give an exposition of some of the topics in mathematical statistics which find application in the representative method of sampling. Necessarily considerable selection of material had to be made. We believe, however, that the problem of the best estimate and an appropriate method of obtaining such an estimate are fundamental, and we hope that our exposition has helped to make clear these concepts of mathematical statistics which have proved so useful in the representative method.

REFERENCES

[1] J. Neyman, "On the two different aspects of the representative method," *Journal of the Royal Statistical Society*, vol. 97 (1934) pp. 558–625.

[2] J. Neyman, *Lectures and conferences on mathematical statistics, Graduate School of the U. S. Department of Agriculture*, (1937) pp. 89–105

[3] J. Neyman, "Contributions to the theory of sampling human populations," *Journal of the American Statistical Association*, vol 33 (1938) pp. 101–116.

[4] S. S. Wilks, *Statistical inference*, Edwards Brothers (1936–37) pp. 8–11, p. 70, pp. 71–86.

[5] R. C. Geary, "The frequency distribution of the quotient of two normal variates," *Journal of the Royal Statistical Society*, vol. 93 (1930) pp. 442–446.

[6] R. A. Fisher, "Inverse probability," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930) p 528.

[7] H. L. Rietz, "A recent advance in statistical inference," *American Mathematical Monthly*, vol. 45 (1938) pp. 149–158.

The University of Iowa.

ON A NEW CLASS OF "CONTAGIOUS" DISTRIBUTIONS, APPLICABLE IN ENTOMOLOGY AND BACTERIOLOGY

By J. Neyman

CONTENTS

1. **Introduction.** There are a number of fields in which experimental data cannot be treated with any success by means of the usual "Student's" test and—very probably—by means of the more general analysis of variance z-test of Fisher. It is known in fact [1] that the t-test, as applied to two samples, is only valid when the populations from which the samples are drawn have equal variances. As the z-test is of a nature similar to the t-test, with the difference that it is applied to detect differentiation in means of more than two populations, a similar conclusion seems very likely. Thus, whenever we have to compare means of populations with distinctly different variances, we have to look for some new tests. It may be useful to mention at once two instances in which the situation mentioned actually arises.

As a first instance we may quote certain entomological experiments. Suppose it is desired to test the efficiency of several treatments intended to destroy certain larvae on a field. The experiments are arranged in the usual way. The treatments compared are applied to particular plots with several replications and then the plots (or smaller parts of them) are inspected and all the surviving larvae are counted Thus the observations represent the numbers of surviving larvae in several equal areas. It happens frequently that, while there is room for doubt as to whether there is any significant difference between the average number of survivors corresponding to particular treatments, there is no doubt whatever that the variability of the observations differs from treatment to treatment.

We have another similar case in bacteriology. The experiments I have in mind consist in determining the bacterial density by the so called "plating method." This consists in taking a number of samples of the analyzed liquid

and in spreading them separately on Petri plates. After a suitable period of time a number of colonies appear on the plates and their numbers represent the observational figures. I am informed that the variability of such observations does not depend very much on the technique of mixing the liquid and of taking the samples—when this technique is on a proper level—but does considerably depend on the kind and on the number of bacteria present in the liquid.

The above examples justify an effort to find some new and more appropriate test. The first step in this direction must consist in an analysis of the machinery behind the observable distributions and in deducing their analytical form. Once this problem is solved and repeated comparisons show a satisfactory agreement between the theory and the observation, we may proceed to the next step and deduce the appropriate tests.

The purpose of the present paper consists in deducing a family of distributions which provide a reasonably good fit in several cases in which they have been tested. It may be hoped that they will prove satisfactory also in many cases in the future.

2. Distribution of larvae in experimental plots. When the problem of the distribution of larvae in experimental plots first arose, attempts were made to fit the Poisson Law of frequency These attempts, however, failed almost invariably with the characteristic feature that, as compared with the Poisson Law, there were too many empty plots and too few plots with only one larva. A similar circumstance is frequently, though not so regularly, observed in counts of microorganisms in single squares of a haemacytometer. These facts suggest that the distributions considered belong to a class which Pólya [3] proposed to call "contagious": the presence of one larva within an experimental plot increases the chance of there being some more larvae. And it is not difficult to see the cause of this dependence. Larvae are hatched from eggs which are being laid in so-called "masses." After being hatched they begin to travel in search of food. Their movements are slow and therefore, whenever in a given plot we find a larva, this means that the mass of eggs, from which it was hatched, must have been laid somewhere near, and this in turn means that we are likely to find in the same plot some more larvae from the same litter. Of course, there may be also others coming from other litters, too.

A similar explanation may apply also to microorganisms counted in single squares of a haemacytometer or to colonies on parallel plates. However, here the situation does not seem as clear as in the case of larvae. As far as the haemacytometer counts are concerned, also another cause of contagiousness may be suggested. Witnessing once the process of preparation of the experiment, I noticed that, immediately after the drop of liquid was deposited into the chamber of the haemacytometer and for some time after, the positions of cells seen under the microscope were not fixed. Some of them seemed to lie on the bottom and the others were floating downwards in an irregular movement. Trying to follow the movements of particular cells I had the impression

that they were slightly attracted by the cells already stationary or semi-stationary on the bottom of the chamber. If this impression of mine is justified, then the attraction of the floating cells by those already on the bottom could explain the contagiousness of the resulting distribution. It is known, however, that this contagiousness is always rather small and that frequently the distribution of cells in the squares of the haemacytometer does follow the Poisson Law very closely.

Owing to the fact that the cause of the contagiousness of the distribution of larvae in experimental plots is clear, we shall deal primarily with the distribution of larvae. Consequently, if the theoretical distributions that we shall deduce fit the empirical ones, we shall be more or less justified in assuming that we guessed the essential features of the actual machinery of movements of the larvae. On the other hand, if the same theoretical distributions appear also to fit satisfactorily the empirical counts of bacteria then in respect of these applications it will be safer to consider that we were lucky enough to find a sufficiently flexible interpolation formula

After these preliminaries we may proceed to a more accurate specification of the conditions of the problem considered. The experimental plot in which the larvae are counted will be denoted by P. We shall make no restriction as to the shape of this plot, but we shall assume that its area, which we shall take as unity, is small compared with that of the experimental field, F. The latter will be assumed to possess M units of area. We shall further assume that the moths laying eggs on the field F select spots for this purpose in a purely random manner. This presupposes that the experimental field is uniform in many relevant respects, e.g. is sown in all its parts by the same kind of plant, etc. Denoting by ξ and η the coordinates of the mass of eggs laid by some particular moth on the field F, we shall treat them as random variables with the elementary probability law

$$p(\xi, \eta) = \frac{1}{M} \tag{1}$$

everywhere within F and zero elsewhere. After the larvae are hatched from the eggs there will be some mortality among them. Let us denote by n the number of larvae hatched from the same mass of eggs, surviving at the moment when the counts are made. We shall treat n as a random variable and denote by $p(n)$ its probability law. At the present moment the writer has no information as to what may be the nature of the function $p(n)$. Consequently it will remain in our calculations in its general form and, wishing to obtain some formulae for immediate calculations, we shall have to substitute for $p(n)$ hypothetical formulae which, on intuitive grounds, may seem plausible If the larvae counted are all more or less of the same age, there is a possibility that $p(n)$ does not differ very much from the Poisson Law, but this point might be verified experimentally and we shall not insist on its being necessarily true.

Consider now a single larva, survivor at the moment of observation, which

was hatched out at a point with coördinates ξ and η. Denote by x and y the coordinates of this larva at the moment of counts. We shall consider x and y as random variables. It is obvious that the probability law of x and y must depend on the values of ξ and η. We shall assume that the dependence is of a particular character; namely, that the probability law of x and y given ξ and η is a function of the differences $x - \xi$ and $y - \eta$. We shall denote it by $f(x - \xi, y - \eta)$.

There is very little that we may consider as known about the function $f(x - \xi, y - \eta)$. It may be treated as describing the habits of travelling of the larvae. There are some indications that there are certain directions in which the larvae tend to travel rather than in others, but they are too vague to be taken into consideration. Only one thing is certain: during the period of time between the birth of the larvae and the moment that the counts are made the larvae are able to travel only at some limited distance. Consequently we shall assume that for sufficiently large values of $|x - \xi|$ and $|y - \eta|$ the function $f(x - \xi, y - \eta)$ is identically zero. Otherwise we shall not make any further assumption concerning $f(x - \xi, y - \eta)$, and it will remain arbitrary in our calculations until we reach the final general formula.

While abstaining from making arbitrary assumptions concerning the habits of single larvae, we shall make one concerning the habits of several of them. This assumption, however, seems to be very plausible. We shall assume that the larvae have no social instincts, so that the random variables x and y corresponding to one larva are independent from those corresponding to any other —*that is to say, apart from the possible dependence on the same pair of* ξ *and* η.

Denote by N the total number of masses of eggs laid on the field F and let k_i be the number of larvae hatched from the i-th mass of eggs, surviving at the moment of observation and present within some particular experimental plot P. Finally let

$$X = \sum_{i=1}^{N} k_i \tag{2}$$

be the total number of larvae to be found within this plot. Our purpose will be to use the above hypotheses in order to determine the probability law of X. In doing so we shall first find that of any of the k_i's. Obviously, when considering just one variable k_i, it would be useless to retain the subscript i, so that below we shall write simply k to denote the number of living larvae, to be found within P, all of which were hatched from the same mass of eggs, situated at some point (ξ, η).

Let us first write the expression for the probability that one particular larva of that group will be found within P. This probability will be a function of ξ and η only, say

$$P(\xi, \eta) = \int\int_P f(x - \xi, y - \eta)\, dx\, dy. \tag{3}$$

Given that the number of survivors of the mass of eggs of the point (ξ, η) is n, the probability that exactly k of them will be found within P will be represented by the binomial formula, say

$$P\{k \mid n, \xi, \eta\} = \frac{n!}{k!\,(n-k)!} P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}. \tag{4}$$

It will be noticed that in writing this formula we use the hypothesis that the larvae have no social instincts.

Multiplying (4) by the probability law of ξ and η, and integrating with respect to those variables over the whole field F, we shall obtain the probability, $P\{k \mid n\}$ that out of the n survivors of a mass of eggs, laid anywhere within F, exactly k larvae will be found within P:

$$P\{k \mid n\} = \frac{n!}{k!\,(n-k)!} \frac{1}{M} \int\int_F P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta. \tag{5}$$

Multiplying this result by $p(n)$ and summing for all values of n, we shall obtain the absolute probability of k having any specified value

However, before doing so, we must use the hypothesis about the function $f(x - \xi, y - \eta)$ to deduce certain consequences concerning the integral in (5).

Originally we did not make any assumption as to the origin of coordinates on the field F. It will be now convenient to assume that it is located somewhere within the experimental plot P, for example in its center or in any other easily specified point. Owing to the particular property of the function $f(x - \xi, y - \eta)$ it will now follow that, for sufficiently large values of ξ and η, the probability $P(\xi, \eta)$ will be equal to zero. Let us denote by A the part of the experimental field where $P(\xi, \eta) > 0$. Obviously A denotes the set of points, a, in F such that, if a mass of eggs is laid in one of them, the distance of a from the plot P is not too large for the larvae hatched in a to reach the plot P before the moment of observation. Obviously also the plot P is included in A. Consequently the area of A, to be denoted by the same letter A, must be greater than unity. Owing to the lack of any precise knowledge of the nature of the function $f(x - \xi, y - \eta)$ it is impossible to say anything about the shape of A.

Let us now turn to the integral in (5). The function under this integral changes its form according to whether the point (ξ, η) is within or without A. If $k = 0$, then the integral in (5) reduces to

$$\int\int_F (1 - P(\xi, \eta))^n\, d\xi\, d\eta = M - A + \int\int_A (1 - P(\xi, \eta))^n\, d\xi\, d\eta. \tag{6}$$

If however $k > 0$, then

$$\int\int_F P^k(\xi, \eta)\,(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta = \int\int_A P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta. \tag{7}$$

Now we can write

$$P\{k\} = \sum_{n \geq 0} p(n) P\{k \mid n\}, \tag{8}$$

which gives in particular

$$P\{k = 0\} = 1 - \frac{A}{M} + \frac{1}{M}\int\int_A \sum_{n\geq 0} (1 - P(\xi, \eta))^n p(n)\, d\xi\, d\eta \tag{9}$$

and for $k > 0$

$$P\{k\} = \frac{1}{M}\int\int_A \sum_{n\geq 0} \frac{n!}{k!\,(n-k)!} P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k} p(n)\, d\xi\, d\eta. \tag{10}$$

This is the general form of the probability law of k, which involves two unspecified functions $p(n)$ and $P(\xi, \eta)$. We shall not analyze it but proceed to the calculation of the characteristic function $\phi_k(t)$ of k, which will then be used to calculate that of X. We have

$$\phi_k(t) = \sum_{k\geq 0} e^{itk} P\{k\} \tag{11}$$

or, using (9) and (10), and after easy transformations

$$\phi_k(t) = 1 - \frac{A}{M}\left(1 - \frac{1}{A}\int\int_A \sum_{n\geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right). \tag{12}$$

Owing to the assumption that the larvae have no social instincts all the variables $k_1, k_2, \cdots k_N$ in (2) must be considered as mutually independent. As the characteristic function of any of them has the same form (12), the characteristic function, $\phi_X(t)$, of their sum, X, will be represented by the Nth power of the expression (12) Denoting by m the average number of masses of eggs per unit of area of the field F, so that $N = Mm$, we shall have

$$\begin{aligned}\phi_X(t) &= \phi_k^N(t) \\ &= \left\{1 - \frac{A}{M}\left(1 - \frac{1}{A}\int\int_A \sum_{n\geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right)\right\}^{Mm}.\end{aligned} \tag{13}$$

This will be the characteristic function of X for any value of M If it is desired to put into effect the assumption that "M is large", we shall have to consider the limit of (13) for $M \to \infty$. This will be denoted by $\phi(t)$ and we shall have

$$\phi(t) = \exp\left\{-Am\left(1 - \frac{1}{A}\int\int_A \sum_{n\geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right)\right\}. \tag{14}$$

In order to obtain the numerical value of the probability of X having any specified value X', it remains only to specify the functions $p(n)$ and $P(\xi, \eta)$ and to use the familiar formula

$$P\{X = X'\} = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \phi(t) e^{-itX'}\, dt. \tag{15}$$

3. **Particular classes of the limiting distribution of X.** Until we have some experimental evidence as to what might be the nature of the two functions $p(n)$ and $f(x - \xi, y - \eta)$ or $P(\xi, \eta)$, we may try a few guesses. If the results

obtained in this way agree with empirical distributions, we shall have some reason to think that the guesses are not altogether wrong

In certain cases all the larvae considered are at the moment of observation approximately of the same age. Alternatively, we may count only larvae which are at the same stage of development. With such counts it is not unreasonable to try for $p(n)$ either the binomial or the Poisson formula. Either of them will lead to easy calculations of (14) Writing

$$p(n) = e^{-\lambda}\frac{\lambda^n}{n!} \tag{16}$$

with λ representing the average number of survivors at the moment of observation per unit mass of eggs, we shall get for $\phi(t)$ the following expression

$$\phi(t) = \exp\left\{-Am\left(1 - \frac{1}{A}\int\int_A e^{\lambda P(\xi,\eta)(e^{it}-1)}\,d\xi\,d\eta\right)\right\}. \tag{17}$$

Substituting here for $P(\xi, \eta)$ any suitable function we shall obtain a corresponding particular form of the characteristic function $\phi(t)$, so that (17) determines a whole family of distributions. Substituting in (14) instead of (16), say the binomial formula, we shall obtain another family of contagious distributions.

Strictly speaking, in order to obtain some particular distribution from the formula (17), we have to specify the function $f(x - \xi, y - \eta)$, then to calculate $P(\xi, \eta)$ and substitute it in (17). Since however we have no knowledge of the properties of $f(x - \xi, y - \eta)$ and have to select it only on intuitive grounds, we may as well select the function $P(\xi, \eta)$. It may be selected either by itself directly, in which case there will be no difficulty in substituting it in (17), or by some indirect method In the other case we may find it more convenient to use another form of (17) which is obtained by expanding the exponential under the sign of the integral in (17) and by integrating term by term, which is obviously permissible. In this way we get

$$\log\phi(t) = Am\sum_{n=1}^{\infty}\frac{\lambda^n(e^{it}-1)^n}{n!}P_n. \tag{18}$$

Where P_n stands for the expression

$$P_n = \frac{1}{A}\int\int_A P^n(\xi,\eta)\,d\xi\,d\eta \tag{19}$$

and has the form of a moment of nth order of a certain probability law which it is easy to determine

We may consider for a moment the value of $P(\xi, \eta)$ as a random variable Z. Its values cannot exceed the limits, zero and unity. Let z be any number between zero and unity and denote by $AF(z)$ the measure of the set of points belonging to A where $P(\xi, \eta) \leq z$. Then the function $F(z)$ will possess all the properties of the integral probability law of a variable Z which we may identify

with $P(\xi, \eta)$ and the integrals P_n will be simply the moments of Z namely, $P_n = \int_0^1 z^n dF$, where, of course, the integral would be considered in the sense of Stieltjes. It is interesting to notice that P_1 is always equal to A^{-1}. To see this consider the integral

$$AP_1 = \int\int_A P(\xi, \eta)\, d\xi\, d\eta \tag{20}$$

and substitute in it the expression of $P(\xi, \eta)$ in terms of the function $f(x - \xi, y - \eta)$. We get

$$AP_1 = \int\int_A d\xi\, d\eta \int\int_P f(x - \xi, y - \eta)\, dx\, dy \tag{21}$$

$$= \int\int\int\int_W f(x - \xi, y - \eta)\, dx\, dy\, d\xi\, d\eta. \tag{22}$$

Where the four-dimensional region of integration W is defined as follows. (i) The variables x and y vary so that the point having them for its coördinates may have any position within, but cannot be outside, of the experimental plot P. (ii) When x and y are fixed in the above way, say $x = x'$ and $y = y'$, then ξ and η may assume all those values for which the function $f(x' - \xi, y' - \eta)$ is positive. Let us denote this system of values of ξ and η by $B(x', y')$. Then we can calculate AP_1 as follows

$$AP_1 = \int\int_P dx\, dy \int\int_{B(x,y)} f(x - \xi, y - \eta)\, d\xi\, d\eta. \tag{23}$$

Now it is easy to see that the second integral in (23) is always equal to unity, whatever be x and y satisfying (i). To see this we have to recall the fundamental property of the function $f(x - \xi, y - \eta)$, due to the fact that it is the elementary probability law of x and y, namely that if ξ and η are fixed in one way or another, and it is integrated with respect to the other pair of variables, over all their values for which it is positive, the result will be equal to unity. In particular we shall have

$$\int\int_{f>0} f(u, v)\, du\, dv = 1. \tag{24}$$

Consider now the second integral in (23) and make the substitution

$$\xi = x - u, \qquad \eta = y - v \tag{25}$$

so that, instead of ξ and η we shall now integrate for u and v It will be seen that the result of this substitution is exactly the integral (24), equal to unity. Since it was assumed that the area of P is equal to unity, it follows that $AP_1 = 1$. This equality is thus the necessary condition that the function $P(\xi, \eta)$ must satisfy. Besides, being a probability, it cannot be negative and cannot exceed unity. Whether any function having these properties may play the rôle of

$P(\xi, \eta)$ must be left for further inquiry. Assuming temporarily that this is so we can tentatively specify the probability laws belonging to the class determined by (18) by substituting in (18) instead of the P_n's the corresponding moments M_n of any distribution function $F(z)$ with its range between zero and unity, remembering only the interpretation of its first moment that we have found above, namely $M_1 = P_1 = A^{-1}$.

4. **Certain general properties of the distributions deduced.** Using the above result, we may substitute it in the formula (18) and get

$$\log \phi(t) = m\lambda(e^{it} - 1) + Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n. \tag{26}$$

Owing to the fact that the first term in the right hand side, $m\lambda(e^{it} - 1)$, represents the logarithm of the characteristic function of the Poisson Law,

$$p(x) = e^{-m\lambda} \frac{(m\lambda)^x}{x!} \tag{27}$$

for $x = 0, 1, 2, \cdots$ the formula (26) is especially interesting. Comparing the formulae

$$\begin{cases} P_1 = \int_0^1 z\, dF = A^{-1} \\ P_n = \int_0^1 z^n\, dF \end{cases} \tag{28}$$

we see that $0 < P_n \leq A^{-1}$ so that $AP_n \leq 1$. This circumstance assures the absolute and uniform convergence of (26). Frequently the higher moments P_n will be much smaller than the first, P_1, and if this tends to zero, all the products AP_n for $n \geq 2$ will do so too. In those cases $\log \phi(t)$ will tend to $m\lambda(e^{it} - 1)$ uniformly for all values of t. To see this take an arbitrary $\epsilon > 0$ and select N so large that

$$m \sum_{n=N+1}^{\infty} \frac{(2\lambda)^n}{n!} < \frac{\epsilon}{2}. \tag{29}$$

Next let A_0 be large enough for

$$AP_n < \frac{\epsilon}{2m} e^{-2\lambda} \tag{30}$$

for all $n = 2, 3, \cdots N$ and for any $A \geq A_0$. For such values of A we shall have

$$\begin{aligned} &\left| Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)}{n!} P_n \right| \\ &\leq Am \left(\sum_{n=2}^{N} \frac{\lambda^n |e^{it} - 1|^n}{n!} P_n + \sum_{n=N+1}^{\infty} \frac{\lambda^n |e^{it} - 1|^n}{n!} P_n \right) < \epsilon \end{aligned} \tag{31}$$

independently of what is the value of t. This result may be formulated as

PROPOSITION I. *If the parameters m and λ remain constant but the probability law $F(z)$ is changed so that all the products AP_n tend to zero for $n = 2, 3, \cdots$, then $\phi(t)$ tends to $m\lambda(e^{it} - 1)$ uniformly for all values of t and, consequently, the corresponding probability law of X tends to that of Poisson, given by* (27).

The above proposition may be considered as an explanation of the circumstance that occasionally the distribution of larvae may be very close to that of Poisson. This may happen for instance when the larvae that we count are sufficiently old and have had a sufficient time to travel very far from the spot where they were hatched. In such cases A will be large and, if the function $f(x - \xi, y - \eta)$ has some appropriate properties, all the products AP_n may be very small. But it is interesting to notice that there is a possibility of A increasing without the products AP_n tending to zero. Such will be for instance the case if $P(\xi, \eta)$ could have within A only two values $B_1(A)$ and $B_2(A)$ changing with A, one close to unity and the other close to zero If Ap and Aq are the areas of the parts of A where $P(\xi, \eta)$ has those two different values, then we shall have

$$\begin{cases} P_1 = pB_1(A) + qB_2(A) = A^{-1} \\ P_n = pB_1^n(A) + qB_2^n(A) \end{cases} \tag{32}$$

and

$$AP_n = \frac{pB_1^n(A) + qB_2^n(A)}{pB_1(A) + qB_2(A)} \tag{33}$$

may tend to unity as A is increased. In such cases the probability law of X will not tend to (27). While calling attention to this possibility, it should be emphasized that it is not likely to occur in practice. In the cases of discontinuous $F(z)$ considered below $P\{X\}$ does tend to (17). The same is true also in such cases where it is assumed that

$$\begin{aligned} \frac{dF}{dz} &= a + bz \geq 0 \qquad \text{for} \quad 0 < 2 < c \leq 1 \\ &= 0 \qquad\qquad\quad \text{elsewhere} \end{aligned} \tag{34}$$

etc.

Before proceeding to specialize the expression (26) of the logarithm of the characteristic function, we shall show the connection existing between the P_n's and the semi-invariants of X. To calculate the latter it is sufficient to differentiate (26) with respect to t, to put $t = 0$, and to divide the result by the appropriate power of i. Denoting by γ_k the kth semi-invariant, by μ_1' the first moment about zero, and by μ_k the kth central moment of X we easily get

$$\begin{cases} \mu_1' = \gamma_1 = m\lambda \\ \mu_2 = \gamma_2 = m\lambda(1 + A\lambda P_2) \\ \mu_3 = \gamma_3 = m\lambda(1 + 3A\lambda P_2 + A\lambda^2 P_3) \\ \mu_4 - 3\mu_2^2 = \gamma_4 = m\lambda(1 + 7A\lambda P_2 + 6A\lambda^2 P_3 + A\lambda^3 P_4) \\ \text{etc.} \end{cases} \tag{35}$$

It will be seen that, in general, the kth semi-invariant depends on P_2, P_3, $\cdots P_k$ only. Another property of the new distributions that we shall mention is that they are "stable"

PROPOSITION II. *If X_1, X_2, $\cdots X_s$ are s independent random variables all following the same distribution with the logarithm of the characteristic function given by* (26), *then the sum $Y = \sum_{i=1}^{s} X_i$ will follow the same probability law with the exception that instead of the parameter m it will depend on the product sm.*

In order to establish this proposition it is sufficient to notice that the logarithm of the characteristic function of the variable Y is equal to the expression (26) multiplied by s.

Lastly, it may be noticed that the family of distributions determined by (26) is different from the comparable distributions deduced by Pólya ([3], p. 153, formulae (40) and (41)). In fact the logarithms of the characteristic functions of the latter could be written as follows:

$$-a \log (1 - b(e^{it} - 1)) = ab(e^{it} - 1) + a \sum_{n=2}^{\infty} \frac{b^n(e^{it} - 1)^n}{n} \tag{36}$$

and

$$\frac{c(e^{it} - 1)}{1 - de^{it}} = \frac{c(e^{it} - 1)}{1 - d} + \frac{c}{1 - d} \sum_{n=2}^{\infty} \left(\frac{d}{1 - d}\right)^{n-1} (e^{it} - 1)^n \tag{37}$$

respectively and, even if the formal expansions in powers of $(e^{it} - 1)$ converge, the identification of those expansions with (26) would require that P_n possess values exceeding unity, which is inconsistent with their essential property of being successive moments of a positive variable $0 \leq Z \leq 1$. Of course, the convergence of (36) and (37) would impose special restrictions on the constants that those formulae involve.

5. **Contagious distribution of type A depending on two parameters.** The simplest assumption that we can make concerning the function $P(\xi, \eta)$ is that it possesses some constant positive value within A and is zero elsewhere. Owing to (20) this constant value must be equal to A^{-1}. Substituting this in (17) we immediately obtain, say

$$\phi_1(t) = \exp\left\{-Am\left[1 - \exp\left(\frac{\lambda}{A}(e^{it} - 1)\right)\right]\right\}. \tag{38}$$

We could use the above formula directly to obtain the corresponding probability law. But before doing so, it may be useful to illustrate the machinery of the alternative method of obtaining the characteristic function of X and to calculate the same formula using (26)

If $P(\xi, \eta)$ is equal to A^{-1} everywhere in A, this means that the function $F(z)$ is a step function, which is equal to zero for any $z < A^{-1}$ and is equal to unity

elsewhere. Accordingly we shall have $M_n = A^{-n}$. Substituting this into (26) instead of P_n we easily get

$$\log \phi_1(t) = Am\left(e^{\frac{\lambda}{A}(e^{it}-1)} - 1\right) \tag{39}$$

which is equivalent with (38).

We shall now proceed to the calculation of the probabilities $P\{x = k\}$ as determined by either (38) or (39). For this purpose it will be useful to notice that the characteristic function (38) depends really on two parameters only, which we shall denote by m_1 and m_2,

$$m_1 = Am, \qquad m_2 = \lambda/A \tag{40}$$

In order to simplify the printing we shall further denote

$$z = m_1 e^{-m_2} \tag{41}$$

Expanding the two first exponentials of the three involved in (38), we may write

$$\phi_1(t) = e^{-m_1} \sum_{k=0}^{\infty} \frac{m_2^k}{k!} e^{ikt} \sum_{n=0}^{\infty} \frac{z^n}{n!} n^k. \tag{42}$$

This is the form of the characteristic function which is the most convenient when we have in mind applying the formula (15). In fact, it will be seen that we may multiply (42) by $e^{-iX't}$ and then integrate the series term by term. Further, it will be noticed that, on integrating between the limits $-\pi$ and $+\pi$, all the terms of the product will vanish except for the one which is independent of t. Consequently, the result of substituting (42) in the right hand side of (15) will be the coefficient of $e^{iX't}$ in the expansion (42), so that

$$P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \sum_{i=0}^{\infty} \frac{z^i}{i!} i^k. \tag{43}$$

As it is easy to verify, we have

$$P\{x = 0\} = e^{-m_1(1-e^{-m_2})} \tag{44}$$

and, for $k \geq 1$

$$P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \frac{d^k}{du^k} e^{m_1 e^{u-m_2}}\Big|_{u=0}. \tag{45}$$

This formula gives an easy check of the identity $\sum_{n=0}^{\infty} P\{x = n\} = 1$. In fact, the left hand side can be looked upon as a product of e^{-m_1} by the Taylor's expansion of the function differentiated in (45) taken at the point $u = m_2$, which gives identically unity.

Successive differentiations give in turn

$$P\{X = 1\} = e^{-m_1(1-e^{-m_2})} \frac{m_2}{1!} m_1 e^{-m_2} \tag{46}$$

$$P\{X = 2\} = e^{-m_1(1-e^{-m_2})} \frac{m_2^2}{2!} (m_1^2 e^{-2m_2} + m_1 e^{-m_2}) \tag{47}$$

etc. Comparing the formulae (44), (46) and (47), the effect of the "contagiousness" of the distribution is easily seen. $P\{x = 2\}$ differs from what it would have been, if the distribution was that of Poisson, by the additional term $m_1 e^{-m_2}$ within the brackets.

Formulae (44), (46) and (47), and others which could be obtained by differentiating as indicated in (45), could be used for numerical calculations. However, these are greatly simplified by the use of the following elegant formula, deduced by Dr. Geoffrey Beall of the Dominion Entomological Experimental Station, Chatham, Ontario.

$$P\{X = n + 1\} = \frac{m_1 m_2 e^{-m_2}}{n+1} \sum_{t=0}^{n} \frac{m_2^t}{t!} P\{X = n - t\}. \tag{48}$$

The correctness of this formula may be easily checked by calculating $P\{X = n - t\}$ from (43) and by substituting it in (48). Simple rearrangements will then give what could be obtained from (43) by putting $k = n + 1$.

Substituting $P_n = A^{-n}$ in formulae (35) and taking account of (40), we get

$$\mu_1' = \lambda m = m_1 m_2 \tag{49}$$

$$\mu_2 = \lambda m \left(1 + \frac{\lambda}{A}\right) = m_1 m_2 (1 + m_2). \tag{50}$$

Solving these equations for m_1 and m_2 we obtain the formulae

$$m_2 = (\mu_2 - \mu_1')/\mu_1', \qquad m_1 = \mu_1'/m_2 \tag{51}$$

If the moments μ_1' and μ_2 are determined for an empirical distribution, these formulae may be used for estimating m_1 and m_2. In cases which were tried, this process did give frequently a satisfactory fit. Sometimes, however, when the tail of the original empirical distribution was very irregular, this distribution was better approximated by calculating the moments μ_1' and μ_2 not from itself but after a certain amount of smoothing of the tail. It follows that the method of fitting the new distribution to the empirical data requires some further study. At present it will suffice to mention that, whenever this distribution was tried on distributions of larvae which at the moment of counts were approximately of the same stage of development, the fit obtained was very satisfactory. It is hoped that a number of actual distributions fitted, together with the description of the method of counting, etc., will be soon published by Dr. Beall. As a matter of illustration one of his distributions is reproduced at the end of the present paper.

As for the distribution considered we have

$$\lim_{A\to\infty} AP_n = \lim A^{-n+1} = 0, \qquad n = 2, 3, \cdots \tag{52}$$

It follows from the above theory that, as $A \to \infty$, the probability law (48) tends to that of Poisson, namely

$$\lim_{A\to\infty} P\{X = n\} = e^{-m_1 m_2} \frac{(m_1 m_2)^n}{n!}. \tag{53}$$

For this reason the distribution (48) could be perhaps called the generalized probability law of Poisson, but it seems that the term "contagious distribution of type A with two parameters" will be more descriptive. Further on we shall see what is the justification of the description "of type A".

It was stated at the outset of the present paper that, when comparing the distributions of larvae in two series of plots subjected to two different treatments, there is sometimes doubt whether the means of those distributions are equal or not, while the difference in variability is more or less obvious. The formulae (49) and (50) give us the explanation of these facts. It is seen from the formula (49) that the mean of the distribution is equal to the product of the mean number of masses of eggs per unit of area and of the mean number of larvae per mass of eggs surviving at the moment of counts. If the two treatments compared are of about the same efficiency of killing the larvae, then the values of λ for each of them will be approximately equal and, consequently, we shall obtain about the same values for the two means. But while being of an equal efficiency as far as the killing is concerned, the two treatments may annoy the larvae in an unequal way. For example if the first treatment is dummy (no treatment) and the other is in general ineffective, it may still spoil the taste of the leaves that the larvae feed on. In such a case they may be compelled to travel a little more than they would otherwise, which will lead to an increase in A Looking at the formula (50), it is easy to see that this would lead to a decrease in the value of μ_2. Alternatively the treatment may produce a temporary paralysis of the larvae which may reduce A and bring an increase of μ_2.

These remarks were applied to moments (49) and (50) of the particular distribution (45), but looking at the formulae (35), it is easily seen that they are true in the general case also.

6. Contagious distributions of type A depending on three parameters. As mentioned before, in order to determine some particular contagious distribution contained in the class depending on equation (18) it is sufficient to substitute in it instead of the P_n the moments of any distribution with its range confined to the interval from zero to unity, with the only restriction that the reciprocal of the first moment should be equal to A. Obviously this could be done in an infinity of ways, all of which will give more or less different results. We shall select the following one, representing a natural generalization of the procedure adopted above and leading to very simple formulae

Formerly we have assumed that $P(\xi, \eta)$ possesses a constant value A^{-1} within the whole area A. At present we may assume that within this area it may possess one of two (three, four, etc.) values, say B_1 and B_2. Considering again $P(\xi, \eta)$ as a random variable Z, this will be equivalent to an assumption that Z may possess only one of the values B_1 and B_2 both positive and not exceeding unity. Again the probabilities of $Z = B_i$ are at our disposal. We shall take that these probabilities are equal, i.e. equal to $\frac{1}{2}$.

Comparing these assumptions with what may be the actual situation, one may be led to think that they are rather artificial. This however is not so. There is no doubt that the value of $P(\xi, \eta)$ does change within A, and it is also probable that the change is smooth. As we have no knowledge of the character of this function we first take its mean value within the area A and treat it as its first approximation. Next we divide the area A into two equal parts, say A_1 and A_2 and so that the greatest value of $P(\xi, \eta)$ in A_1 does not exceed any of the values in A_2. Then taking the average of $P\{\xi, \eta\}$ within A_1 and a similar average within A_2 and denoting them by B_1 and B_2 respectively, we do obtain a better approximation to the actual values of $P(\xi, \eta)$ assuming that it is equal to B_i everywhere in A_i. That is, in fact, the real meaning of the hypothesis formulated above and that we are going to accept in the following.

Denoting again by M_n the moments of Z we shall have

$$M_1 = \tfrac{1}{2}(B_1 + B_2) = A^{-1} \tag{54}$$

and generally

$$M_n = \tfrac{1}{2}(B_1^n + B_2^n). \tag{55}$$

Substituting (55) in (26) we get, say

$$\phi_2(t) = \frac{Am}{2}\,(e^{\lambda B_1(e^{it}-1)} + e^{\lambda B_2(e^{it}-1)} - 2). \tag{56}$$

We notice that this expression depends on three parameters, say

$$m_1 = Am, \qquad m_2 = \lambda B_1, \qquad m_3 = \lambda B_2. \tag{57}$$

In order to get the formulae for the probabilities of X having any specified values we could again apply the method used above when treating the more simple case. It may be useful however to illustrate a shorter way which easily leads to a generalization of Dr. Beall's recurrence formula As we have noticed before, the probability $P\{X = k\}$ is equal to the coefficient of e^{ikt} in the expansion of the characteristic function in powers of e^{it}. Substituting for simplicity $z = e^{it}$, so that $t = -i \log z$, we may say that, if $\phi(t)$ is the characteristic function of a variable X_1 which is able to possess only integer values, then $P\{X = k\}$ is equal to the coefficient of z^k in the expansion of say $\psi(z) = \phi(-i \log z)$ Applying this rule to (56) we can write the following expression for the generating function $\psi(z)$,

$$\psi(z) = e^{-m_1} e^{\frac{1}{2}m_1\{e^{m_2(z-1)} + e^{m_3(z-1)}\}} = \sum_{k=0}^{\infty} z^k P\{X = k\}. \tag{58}$$

In other words

$$P\{X = 0\} = \psi_{(0)} = e^{-m_1} e^{\frac{m_1}{2}\{e^{-m_2} + e^{-m_3}\}} \tag{59}$$

$$P\{X = k\} = \frac{1}{k!} \frac{d^k \psi}{dz^k}\bigg|_{z=0}, \qquad k = 1, 2, \ldots \tag{60}$$

But

$$\frac{d\psi}{dz} = \frac{m_1}{2} \psi(z) \{m_2 e^{m_2(z-1)} + m_3 e^{m_3(z-1)}\} \tag{61}$$

$$= \frac{m_1}{2} \psi(z)\chi(z) \qquad \text{(say)}$$

and it is easy to see that generally

$$\frac{d^k \chi}{dz^k} = m_2^{k+1} e^{m_2(z-1)} + m_3^{k+1} e^{m_3(z-1)}. \tag{62}$$

As the kth derivative of $\psi(z)$ in (60) may be calculated by applying the familiar formula for the $(k-1)$st derivative of the product $\psi(z)\chi(z)$ in (61), we obtain

$$\frac{d^{n+1}\psi}{dz^{n+1}}\bigg|_{z=0} = \frac{m_1}{2} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \left(\frac{d^k \chi}{dz^k} \frac{d^{n-k}\psi}{dz^{n-k}}\right)\bigg|_{z=0}. \tag{63}$$

Using the formulae (60) and (62) we immediately obtain

$$P\{X = n+1\} = \frac{m_1}{2(n+1)} \sum_{k=0}^{n} \frac{m_2^{k+1} e^{-m_2} + m_3^{k+1} e^{-m_3}}{k!} P\{X = n-k\}. \tag{64}$$

As whenever $B_1 = B_2$ and consequently $m_2 = m_3$, the distribution considered now becomes identical with that considered formerly, depending on two parameters only, it is seen that the formula (64) represents a direct generalization of the formula (48). For purposes of successive calculation of the probabilities it will be probably more convenient to write (64) in the following form

$$P\{X = n+1\} = \frac{m_1 m_2 e^{-m_2}}{2(n+1)} \sum_{k=0}^{n} \frac{m_2^k}{k!} P\{X = n-k\} + \frac{m_1 m_3 e^{-m_3}}{2(n+1)} \sum_{k=0}^{n} \frac{m_3^k}{k!} P\{X = n-k\}. \tag{65}$$

This device of finding a recurrence formula for the probabilities will always succeed whenever there are no difficulties in finding the value of the nth derivative of the function χ.

It may be easily shown that if m and λ remain fixed but A tends to infinity, then the distribution (60) tends to the Poisson Law of frequency. Owing to

the general result stated in Proposition I, in order to show this it is only sufficient to prove that for $n \geq 2$

$$\lim_{A\to\infty} AM_n = \lim_{A\to\infty} \frac{B_1^n + B_2^n}{B_1 + B_2} = 0. \tag{66}$$

As both B_1 and B_2 must be included between zero and unity and their sum is equal to $2A^{-1}$, it follows that

$$0 < B_1 \leq A^{-1} \leq B_2 < 2A^{-1}. \tag{67}$$

Therefore

$$0 < AM_n < \frac{1 + 2^n}{2} A^{-n+1} \tag{68}$$

and (66) becomes obvious.

Substituting the values of M_2 and M_3 instead of P_2 and P_3 in the general expressions (35) of the moments, and taking into account the formulae (57), we obtain

$$\begin{cases} \mu_1' = \frac{1}{2}m_1(m_2 + m_3) \\ \mu_2 = \frac{1}{2}m_1(m_2 + m_3 + m_2^2 + m_3^2) \\ \mu_3 = \frac{1}{2}m_1(m_2 + m_3 + 3(m_2^2 + m_3^2) + m_2^3 + m_3^3). \end{cases} \tag{69}$$

If it is desired to fit the distribution to some empirical one using the method of moments, then these formulae could be solved with respect to m_1, m_2 and m_3. We may proceed as follows. Write

$$a = 2\mu_1^1, \qquad b = 2(\mu_2 - \mu_1'), \qquad c = 2(\mu_3 + 3\mu_2 + 2\mu_1'). \tag{70}$$

Then

$$m_1(m_2 + m_3) = a \tag{71}$$

$$m_1(m_2^2 + m_3^2) = b \tag{72}$$

$$m_1(m_2^3 + m_3^3) = c. \tag{73}$$

Multiplying the first of these equations by m_2 and subtracting the result from the second and repeating the same process with the second equation and the third, we get

$$\begin{aligned} m_1m_3(m_3 - m_2) &= b - am_2 \\ m_1m_3^2(m_3 - m_2) &= c - bm_2 \end{aligned} \tag{74}$$

and it follows

$$m_3 = \frac{c - bm_2}{b - am_2} \tag{75}$$

or

$$\frac{b}{a}(m_2 + m_3) - m_2 m_3 = \frac{c}{a}. \tag{76}$$

Again, dividing (73) by (71) we get

$$(m_2 + m_3)^2 - 3m_1 m_2 = \frac{c}{a}. \tag{77}$$

Multiplying (76) by 3 and subtracting from (77), we obtain

$$s^2 - 3bs/a - 2c/a = 0, \tag{78}$$

where $s = m_2 + m_3$. It follows that

$$s = \frac{3b}{2a} \pm \sqrt{\left(\frac{3b}{2a}\right)^2 - \frac{2c}{a}} \tag{79}$$

$$m_2 m_3 = p = \frac{bs - c}{a} \tag{80}$$

$$m_2 = \tfrac{1}{2}(s - \sqrt{s^2 - 4p}) \tag{81}$$

$$m_2 = \tfrac{1}{2}(s + \sqrt{s^2 - 4p}) \tag{82}$$

$$m_1 = a/s. \tag{83}$$

Following these steps we finally arrive to the values of all three parameters, given by the last three formulae.

If the values of the moments μ_1', μ_2 and μ_3 were known without error, the above formulae would give accurate values of m_1, m_2 and m_3. If, however, the moments are estimated from a sample, then the reader must be prepared that, even if the observed variable follows exactly the law, occasionally the sampling errors in the moments will make it impossible to carry out all the calculations indicated. Especially this may easily happen when the true values of m_2 and m_3 are equal or nearly equal, so that the empirical distribution is close to that given by the contagious distribution with only two parameters. As it is seen from (81) and (82), in such a case the true values of s and p must satisfy the relation

$$s^2 - 4p = 0. \tag{84}$$

However, the sampling errors in the moments will ascribe to the left hand side of (84) a value only approximately equal to zero, which may be either positive or negative. In the latter case we shall not be able to use (81) and (82) to estimate m_2 and m_3. As a matter of fact, the above circumstance actually arose in one case when it was tried to fit the three parameters distribution to a set of data which were excellently fitted by a simpler formula (45) involving only two parameters. As mentioned before, the problem of fitting the distributions which are deduced here requires further consideration.

Looking back on the method by which we have substituted a contagious distribution with three parameters m_1, m_2, m_3 for the simpler one with only two parameters, it is easily seen that it can be carried further leading to distributions with four, five, etc. parameters. In each case we would mentally divide the area A in a number of parts of equal size so that the values of $P(\xi, \eta)$ in the first never exceed those in the second, etc. Denoting the average values of $P(\xi, \eta)$ in those areas by $B_1, B_2, \cdots, B_r$, we shall obtain the moments

$$M_n = \frac{1}{r} \sum_{i=1}^{r} B_i^n, \tag{85}$$

substitute them in (26) and proceed more or less as we did above. All the distributions which may be obtained in this way possess certain common traits and I propose to call them "of type A". If the number of parameters in such a distribution is sufficiently high, it seems practically certain that the function $P(\xi, \eta)$ will be well approximated and we may hope to get an excellent fit. However, if a good fit may be attained only by introducing a great number of parameters, it usually means that the method of introducing those parameters is not very successful, and therefore it does not seem worth while to discuss in greater detail the distributions of type A with the number of parameters exceeding three. Instead we shall briefly indicate another class of distributions, built on another principle, which may be called of type B or C.

7. **Contagious distributions of types B and C.** As mentioned before, whenever the distributions of type A were tried on data, the character of which did not obviously contradict the basic assumptions of the theory (approximate equality of age of the larvae), the results were always satisfactory. However, our present experience is rather limited and it is well to anticipate the failures. We may expect that these will be caused by the over-simplified assumptions concerning the function $P(\xi, \eta)$

In order to deal with such a case we may assume that for $0 < z < 1$ the derivative of $F(z)$ exists and is either a linear function of z or is equal to zero. Writing $p(z) = dF/dz$ we shall put

$$\begin{aligned} p_1(z) &= \tfrac{1}{2}A \quad \text{for } 0 < z < 2A^{-1}, \quad A \geq 2 \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{86}$$

Alternatively we may write, say

$$\begin{aligned} p_2(z) &= \frac{2A^2}{9}(3A^{-1} - z) \qquad \text{for } 0 < z < 3A^{-1} \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{87}$$

In the first case we shall obtain, say

$$M_n' = \frac{1}{n+1}\left(\frac{2}{A}\right)^n. \tag{88}$$

On the other hand, the moments of $p_2(z)$ will be given by

$$M''_n = \frac{2(3A^{-1})^n}{(n+1)(n+2)}. \tag{89}$$

Substituting these expressions in (26) we shall easily obtain the two new forms of the characteristic function of X, say

$$\log \phi_3(t) = -m_1 + m_1 \frac{e^{m_2(e^{it}-1)} - 1}{m_2(e^{it}-1)}, \tag{90}$$

with

$$m_1 = Am \quad \text{and} \quad m_2 = 2\lambda/A. \tag{91}$$

Accordingly, the generating function of the probabilities will be, say

$$\psi_3(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)}-1}{m_2(z-1)}} = \sum_{n=0}^{\infty} z^n P\{X = n\}. \tag{92}$$

The distribution determined by (92) may be called of type B.

Using the moments (89) and substituting them in the usual way in (26), we obtain, say

$$\log \phi_4(t) = -m_1 + 2m_1 \frac{e^{m_2(e^{it}-1)} - 1 - m_2(e^{it}-1)}{m_2^2(e^{it}-1)^2}, \tag{93}$$

with

$$m_1 = Am \quad \text{and} \quad m_2 = 3\lambda/A. \tag{94}$$

The probabilities of X having any specified value will be generated by the function, say

$$\psi_4(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)}-1-m_2(z-1)}{m_2^2(z-1)^2}} = \sum_{n=0}^{\infty} z^n P\{X = n\}. \tag{95}$$

The probability law determined by (95) may be called of type C. The comparative merits of all those distributions could be judged by comparing them with the results of observation.

8. **Illustrative Examples and Concluding Remarks.** Any series of positive terms adding up to unity may be considered as determining a probability law of a discontinuous variable such as the X considered above. When trying to obtain probability laws fitting the empirical distributions of some particular origin, the distributions of the numbers of larvae in experimental plots, or the like, we could really start by considering series of some positive terms each depending on one or more parameters, say

$$u_0(m_1, m_2),\ u_1(m_1, m_2),\ u_2(m_1, m_2), \cdots, u_n(m_1, m_2), \cdots \tag{96}$$

and having the property that, whatever the values of those parameters, $\sum_{n=0}^{\infty} u_n(m_1, m_2) = 1$. Studying a considerable number of empirical distributions,

we could apply the "method" of trial and error to guess the form of dependence of the $u_n(m_1, m_2)$ on the m's so that for a broad class of empirical distributions there would be a system of values of the m's, for which the series (96) would satisfactorily fit the data. If we succeed in this task we shall be entitled to a considerable satisfaction as the solution that we obtained would permit various further studies, e.g. the deduction of tests of significance applicable, or approximately applicable, in various cases, and so on.

Looking back at the history of statistics we shall find that the systems of frequency curves of Pearson, of Bruns-Charlier and others belong to the class of results just discussed They are very important—and this especially applies to the Pearson curves—because of the empirical fact, that it is but rarely that we find in practice an empirical distribution, which could not be satisfactorily fitted by any of such curves. Consequently, wishing to deduce some test applicable in this or that case, we may usefully assume that the basic distribution is one of the Pearson system and, owing to the frequently continuous character of the connection between the conditions and the final results, our final formula will be approximately valid when applied to the data under consideration.

This point of view is not unfamiliar in pure mathematics. For example, we know that a broad class of functions may be approximated with any prescribed accuracy by means of polynomials. Wishing to prove a theorem applicable to this class of functions, we sometimes start by proving it for polynomials and then conclude that it is also true for the whole class. Here the rôle of polynomials is perfectly analogous to that of Pearson curves and could be described as that of good interpolation formulae.

But the problem of deducing theoretical distributions could be also considered from a slightly different point of view. Here again we require that the theoretical distribution fits satisfactorily the empirical data. But we may legitimately require something else: an "explanation" of the machinery producing the empirical distributions of a given kind. I have enclosed the word "explanation" in quotation marks so as not to suggest that I am attaching to it too much importance. Mathematics is always dealing with the conceptual sphere which is quite distinct from the perceptual and, at most, admits the possibility of establishing some correspondence. Therefore, however hard we try, we can never produce anything like a real mathematical explanation of any phenomena but instead only some "interpolation formula", some system of conceptions and hypotheses, the consequences of which are approximately similar to the observable facts. But this similarity may be differently placed. In the case of Pearson's curves it applies to the shape of these curves and to the shape of the empirical histograms. Otherwise it may apply to certain real features of the phenomena studied and to some mathematically described model of the same phenomena And if the theoretical distributions deduced from the mathematical model do agree with those that we observe, and if that agreement is more or less permanent, we say that the mathematical model has "explained" the origin of the distributions.

If the problem of deducing interpolation formulae, sufficiently flexible to represent adequately a class of distributions, is of considerable interest, then that of producing similar formulae but involving an "explanation" of the phenomena studied, seems to be still more interesting. Of course, for it to be considered as successfully solved, the theoretical distributions deduced must fit the empirical ones, of a clearly specified kind, "practically always". At the

TABLE I

Distribution of European corn borers in 120 groups of 8 hills each, (data provided by Dr. Beall), fitted by Poisson Law and by type A Law with two parameters

No of borers	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	5.0	24	22.6
1	16.0	16	16.7
2	25.3	16	18.3
3	26.7	18	16.4
4	21.1	15	13.4
5	13.4	9	10.3
6	7.1	6	7.5
7	3.2	5	5.2
8	1.3	3	3.5
9	.4	4	2.3
10	.1	3	1.5
11		0	
12		1	
Beyond		—	2.3
m_1	—	—	2.178
m_2	—	—	1.454
P_{χ^2}	.000,000		.95

TABLE II

Distribution of yeast cells in 400 squares of haemacytometer observed by "Student" (1907), fitted by Poisson Law and by type A Law with two parameters

No. of cells	Frequency		
	Exp P L	Observed	Exp. T A.
0	202	213	214.8
1	138	128	121.3
2	47	37	45.7
3	11	18	13.7
4		3	3.6
5		1	.8
Beyond	2	—	.1
m_1	—	—	3.605
m_2	—	—	.189
P_{χ^2}	$>.02$		$>.1$

present time we may quote a number of instances where it was possible to establish a mathematical probabilistic model of some class of phenomena determining probability laws which fit the empirical distributions with a remarkable accuracy. Perhaps the most important class of these phenomena is provided by the Mendelian theory; a number of other examples, although of a lesser importance but still interesting, have been mentioned elsewhere [2]. In all of them success-

ful checks and rechecks increase our confidence that the conclusions based on the mathematical model determining the theoretical distributions will satisfactorily apply to observational data and also that our interpretations of various constants is more or less correct.

Now, what is the situation with the contagious distributions deduced above? They do represent an attempt to give good interpolation formulae involving an "explanation" of the observable phenomena, and all the constants introduced have meanings which are easy to interpret Owing to the fact that in the process of the larvae surviving and spreading over the field there are certain unknown features, the final general formula that we have deduced involves two arbitrary functions $p(n)$ and $P(\xi, \eta)$. By substituting for them any appropriate functions that the intuition may suggest, we can obtain a number of distributions, each of which may or may not provide a satisfactory interpolation formula. Whether they do or not, must be empirically tested.

Up to the present time the contagious distributions of type A were tried on 12 distributions of larvae and on three distributions of yeast cells in squares of the haemacytometer, which did not quite agree with the Poisson Laws. The results of these trials were always the same: The type A distribution with two parameters provided an excellent fit, which was never worse than that of the more elaborate distribution with three parameters. This circumstance seems encouraging, but future experience may be less satisfactory and it would be very desirable to have some more empirical distributions and checks.

The following table gives two empirical distributions fitted with Poisson Law and with its generalization, as provided by the type A distribution with two parameters.

REFERENCES

[1] P. L. Hsu, "Contribution to the Theory of 'Student's' t-Test as Applied to the Problem of Two Samples." *Statistical Research Memoirs*, Vol. II, (1938) pp. 1-24.

[2] J. Neyman: "Lectures and Conferences on Mathematical Statistics," published by the Graduate School of the United States Department of Agriculture, Washington, D. C., 1938.

[3] G. Pólya: "Sur quelques points de la théorie des probabilités." *Annales de l'Institut Henri Poincaré*, Vol 1, (1931) pp. 117-162.

[4] "Student": "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. 5, (1907), pp. 351-364.

University of California,
Berkeley, California.

ON CONFIDENCE LIMITS AND SUFFICIENCY, WITH PARTICULAR REFERENCE TO PARAMETERS OF LOCATION

By B. L. Welch

1. **Introduction.** The solution of the problem of estimating an interval in which a population parameter should lie, by means of what is now often termed the fiducial type of argument, dates back to the early writers on the theory of errors. However, owing to their lack of "Student's" z distribution, their statements were usually only of an approximate character, and, furthermore, the logical distinction between the fiducial method and the method of inverse probability was never clearly drawn, before R. A. Fisher discussed the subject. It is of interest to note how far "Student" himself went in this matter. In describing the tables which he gave in his original paper he says:[1]

"The tables give the probability that the value of the mean, measured from the mean of the population, in terms of the standard deviation of the sample, will lie between $-\infty$ and z. Thus, to take the tables for samples of six, *the probability of the mean of the population lying between* $-\infty$ *and once the standard deviation of the sample is* 0.9622 or the odds are about 24 to 1 that the mean of the population lies between these limits. The probability is therefore 0.0378 that it is greater than once the standard deviation, and 0.0756 that it lies outside $\pm$ 1.0 times the standard deviation."

It should be noted that "Student's" z is $(\bar{x} - \theta)/s$ where θ is the true population mean. His tables tell us that for $n = 6$, $P(z < 1)$[2] is equal to 0.9622. Owing to the symmetry of the z distribution this is equivalent to saying that $P(z > -1)$ is 0.9622, i.e.

$$P\left\{\frac{\bar{x} - \theta}{s} > -1\right\} = 0.9622.$$

This may be transposed to read

$$P\{\theta < \bar{x} + s\} = 0.9622 \tag{1}$$

which is the statement I have italicized in the above extract, it being there understood that the mean of the population is being measured from the mean of the sample. "Student" therefore makes here what is now called a fiducial statement. In the next sentence he, in effect, attaches a probability to an interval estimate for the population mean. In doing this "Student" was not conscious of introducing any new principle, nor does he apply the method consistently

[1] "Student" (1908). "The Probable Error of a Mean." *Biometrika* VI, p. 20.

[2] P is used to denote the probability of the truth of the relation in the bracket following.

to other problems of estimation. For instance, in discussing the estimation of the correlation coefficient ρ about the same time, he formulates the problem in terms of inverse probability, although he was fully aware of the difficulties involved in postulating an *a priori* distribution for ρ.

In discussing the problem of interval estimation more generally, I shall adopt some of the terminology used by J. Neyman.[3] The sample observations $x_1, x_2, \cdots x_n$ will be noted collectively by E (standing for the "event" point when the observations are represented as coördinates in a space of n dimensions). Then if θ is an unknown parameter, α a fixed probability, and $F(E, \theta, \alpha)$ a function such that

$$P\{F(E, \theta, \alpha) > 0\} = \alpha \tag{2}$$

we may obtain an interval estimate for θ as follows. Let $\delta(E, \alpha)$ denote the set of values of θ such that for any θ in the set we have $F(E, \theta, \alpha) > 0$. Then if we use the notation $\{\delta(E, \alpha) \subset \theta\}$ to indicate that the set $\delta(E, \alpha)$ contains or "covers" the true parameter θ we shall be able to rewrite (2)

$$P\{\delta(E, \alpha) \subset \theta\} = \alpha. \tag{3}$$

We can then adopt the following rule to obtain an interval estimate for θ: (a) calculate from the sample the set $\delta(E, \alpha)$, (b) make the statement that $\delta(E, \alpha)$ covers θ. In adopting this rule we shall be right in the proportion α of cases.

There are, in general, an infinite number of ways in which we can start with a statement of the type (2) to reach the statement of type (3). Neyman has discussed methods of making the best choice between such statements. His approach to this problem may be illustrated by the following example.

Suppose we have a random sample of n from a normal population with standard deviation σ and let

$$s^2 = \frac{\sum (x - \bar{x})^2}{(n-1)},$$

and w = range = largest x − smallest x.
Then we can find a constant b_α such that

$$P\left\{\frac{s}{\sigma} > b_\alpha\right\} = \alpha \tag{4}$$

and, turning this round, we obtain

$$P\left\{\sigma < \frac{s}{b_\alpha}\right\} = \alpha. \tag{5}$$

This means that, if we choose $\alpha = .99$ (say), then we can say that σ is less than $s/b_{.99}$ and in 99% of cases we shall be correct in this statement.

[3] J. Neyman (1937). "Outline of a theory of statistical estimation based on the classical theory of probability." *Phil. Trans. Roy. Soc* A 236, pp. 333–380.

Now similarly we can find c_α such that

$$P\left(\frac{w}{\sigma} > c_\alpha\right) = \alpha \tag{6}$$

and reversing this

$$P\left(\sigma < \frac{w}{c_\alpha}\right) = \alpha. \tag{7}$$

This statement is not inconsistent with (5). It means that, if we choose to base our rule of estimation *always* on the range, then in 99% of cases we shall be correct in saying that $\sigma < w/c_{.99}$. On the other hand, (5) relates to the consequences of applying *always* a rule of estimation based on the standard deviation of the sample. Both (5) and (7) are in themselves true statements, but we must decide which of them is the better one to use. In certain circumstances speed of calculation may be the determining factor, in which case (7) may be preferable, but here we shall assume that the time spent on calculation is not important.

In making the statement that σ is less than some upper limit which is a function of the sample observations, we shall, in general, prefer that this upper limit be placed as low as possible consistent with the chosen confidence coefficient α. We find, however, that it is not possible to say that, whatever the sample obtained, s/b_α will be less than w/c_α or *vice versa*. We must, therefore, approach the problem from another angle. If σ' is a value greater than the true standard deviation σ we can theoretically evaluate the *probability* that $\sigma' < s/b_\alpha$, and similarly the probability that $\sigma' < w/c_\alpha$. We may now express our general desire to place the upper confidence limit for σ as low as possible in a more concrete form. We may ask that the probability that σ' is less than this limit should be as small as possible. We find in the present problem that, whatever $\sigma' > \sigma$, we should include σ' in the interval from 0 to s/b_α less frequently than we should in an interval based on any other statistic. This constitutes an argument for using s rather than any other statistic such as w.

In general, Neyman makes all problems of choosing between alternative procedures of interval estimation depend on the probability that the intervals include values of the parameter different from the true value, as well as on the probability of them containing the true value. This principle of choice does, I think, appear reasonable, although its application is not, of course, so straightforward when statistics with properties of sufficiency similar to those of s do not exist. It is then necessary to introduce other conditions into the formulation of the problem. I intend to discuss elsewhere ways in which this has been done.

To summarize, we may say: (a) we can make many true statements of the type (3); and (b) if we can agree on certain further properties which these statements should possess, we can choose which is the best statement of this type to adopt as our general rule for interval estimation. There are certain differences

between this approach and that of R. A. Fisher, whose attitude is expressed clearly in his contribution to the discussion following Neyman's paper[4] "On the two different aspects of the representative method." Fisher says there that: "In particular he would apply the fiducial method, or rather would claim unique validity for its results, only in those cases for which the problem of estimation proper had been completely solved, i.e. either when there existed a statistic of the kind called *sufficient*, which in itself contained the whole of the information supplied by the data, or when, though there was no sufficient statistic, yet the whole of the information could be utilized in the form of *ancillary* information." Thus it appears that when sufficient statistics do not exist, excepting in those further cases where Fisher claims that the problem of estimation has been completely solved, he would definitely discourage the use of the fiducial argument at all. Neyman, on the other hand, would allow the attempt to obtain interval estimates on the lines described above. Where sufficient statistics do exist, the two approaches do not lead to any final disagreement. Neyman, using results obtained in the Neyman-Pearson theory of testing hypotheses, is led to criteria depending in a particular way on the joint probability law of the sample, and these criteria are seen to involve the sample values only through statistics which have been defined as sufficient. One may regard this fact in two ways: (a) one may say that because a certain line of approach, which seems intuitionally sound, leads to the use of statistics which have been defined as sufficient, therefore this definition of sufficiency is a good one, or (b) one may say that the definition of sufficient statistics is fundamental, and that any method of approach which leads to their use has thereby obtained some extra support.

There remains the case alluded to above, where the joint probability law of the sample does not depend on the unknown parameter θ by way of one statistic only, but where nevertheless it has been said that the problem of estimation has been completely solved. This case will be discussed in the next section.

2. Interval Estimates of Location. R. A. Fisher has given, as a particular example, a case where the unknown parameter is one of location, so that we can write

$$p(x \mid \theta) = \phi(x - \theta).$$

Now if we have a sample of n from this distribution, the $(n - 1)$ differences between successive observations when arranged in order of magnitude will have a joint distribution independent of θ. Hence if we denote the sample by E, and the $(n - 1)$ differences jointly by C, we have

$$p(E \mid \theta) = p(T \mid C, \theta)p(C) \tag{8}$$

where T is some statistic, such as the mean or median, whose distribution does depend on θ and may hence be taken as an estimate of θ. We may therefore

[4] J. Neyman (1934). *J. R. Statist. Soc.* 97, p. 617.

read (8) as follows: the joint probability law of the sample is equal to the probability law of the estimate in samples of the same configuration, C, multiplied by the probability of the configuration, the latter not depending on the unknown θ. From this it has been deduced that all the information respecting θ provided by the sample is given by referring T to the distribution $p(T \mid C, \theta)$. Fisher,[5] for instance, says that "in interpreting our estimate (we) may take as its sampling distribution that appropriate to only those samples which have the actual configuration observed." Later in the same context he remarks that in general, when θ is a parameter of any type whatever, and not necessarily one of location or scaling, if something can be found "corresponding with the configuration of the sample in the simple case discussed above, . . . one of the primary problems of uncertain inference will have reached its complete solution. If not, there must remain some further puzzles to unravel."

It is clear, therefore, that more has been claimed for this method than that it is *practically* useful, or that it yields the best results possible in *large* samples, or that it yields results *highly approximating* the best possible in small samples. There is an emphasis here on completeness that leads one to suppose that all problems of estimation and testing hypotheses may be answered to the best advantage by considering only the distribution of an estimate in samples of the same configuration, the estimate thus attaining properties analogous to those of a sufficient statistic. That this supposition is not true may be seen by considering the following simple example. This example concerns the simplest situation that one deals with in the theory of testing statistical hypotheses. Its relevance to the problem of interval estimation will, however, not be difficult to see.

Suppose that we have a sample from a population involving only a parameter of location θ, and that we wish to test whether θ is equal to θ_0 (say), and that besides θ_0 there is only one value θ_1 (say) which it is possible for θ to take Suppose we require to set up a statistical test which will reject the hypothesis $\theta = \theta_0$, in only a small proportion ϵ of cases, when it is true. Many such tests are possible, and it is natural to choose from them that test which will lead most frequently to the rejection of the hypothesis that $\theta = \theta_0$ when the single alternative $\theta = \theta_1$ is true. Neyman and Pearson[6] have shown that the best test from this viewpoint is provided by the criterion

$$J = \frac{p(E \mid \theta_1)}{p(E \mid \theta_0)}. \tag{9}$$

This criterion must be referred to its distribution in *all* samples when $\theta = \theta_0$. We must therefore choose a constant J_ϵ such that

$$P(J > J_\epsilon \mid \theta = \theta_0) = \epsilon \tag{10}$$

[5] Fisher, R. A. (1936). "Uncertain Inference." *Proc Amer. Acad. Arts and Sciences*, 71, No. 4, p. 257.

[6] J. Neyman and E. S. Pearson (1932). "On the problem of the most efficient tests of statistical hypotheses" *Phil. Trans. Roy. Soc.* A 231, p. 300.

and reject the hypothesis that $\theta = \theta_0$ when $J > J_\epsilon$. This is known to be the best test in these circumstances, and we may demand that any other procedure which claims to use the data exhaustively should be equivalent to it. Now if we decide to use only the distribution of the statistic T in samples of the same configuration, we are led to take as the most powerful test based on $T \mid C$ one which would reject the hypothesis that $\theta = \theta_0$ when the ratio of $p(T \mid C, \theta_1)$ to $p(T \mid C, \theta_0)$ exceeds a certain value. Now by (8) this ratio is exactly the criterion J of (9) above. There is, however, this difference, that J has now to be referred to its distribution in samples with the *same configuration* C as that observed. We shall therefore have to choose $J_\epsilon(C)$ such that

$$P\{J > J_\epsilon(C) \mid C, \theta\} = \epsilon. \tag{11}$$

A test, then, which rejects the hypothesis that $\theta = \theta_0$ when $J > J_\epsilon(C)$ will be such that it is the most powerful possible with respect to the alternative $\theta = \theta_1$, based on samples with the same configuration. However, in actual sampling from a population, we derive samples with all configurations, and the real power of the test will therefore be measured by

$$P\{J > J_\epsilon(C) \mid \theta_1\} = \int P\{J > J_\epsilon(C) \mid C, \theta_1\} p(C)\, dC. \tag{12}$$

This quantity cannot be greater, and will in general be less, than the power[7] of the other test, viz $P(J > J_\epsilon \mid \theta_1)$. (If $J_\epsilon(C)$ is the same for all C, and therefore equal to J_ϵ, the powers will be equal This will be the case when there is a sufficient statistic for θ.) We must therefore conclude that, in relation to this simple problem at least, a method which takes account only of distributions in samples with the same configuration will not use the data to the best advantage.

Of course the type of problem to be solved is usually not so straightforward as the present one. There will usually be more than one value of θ alternative to θ_0, and no uniformly most powerful test will, in general, exist. It is legitimate, however, to consider the above example, because any procedure claiming properties of sufficiency should be able to deal with it in the best possible way.

An example may make the above points clearer, and will show their relevance to the problem of interval estimation. Consider a rectangular distribution with mean θ, and range from $(\theta - \frac{1}{2})$ to $(\theta + \frac{1}{2})$. Let x_1 and x_2 be a sample of 2 from this population, and suppose we require confidence limits for θ such that the chance of them enclosing θ is α.

If we represent x_1 and x_2 as coordinates of a point with respect to rectangular axes, the joint probability distribution is constant over a square centered at the point (θ, θ). This is shown by $ABCD$ in Fig. 1. We have

$$p(x_1, x_2)dx_1dx_2 = dx_1dx_2 \begin{cases} \theta - \frac{1}{2} < x_1 < \theta + \frac{1}{2} \\ \theta - \frac{1}{2} < x_2 < \theta + \frac{1}{2}. \end{cases} \tag{13}$$

[7] Power is used throughout in the Neyman-Pearson sense, i.e. to denote the chance of a test rejecting a hypothesis when a given alternative is actually true.

If we write $z_1 = \frac{1}{2}(x_1 + x_2)$; $z_2 = \frac{1}{2}(x_1 - x_2)$, z_2 will represent the configuratic of the sample, and z_1 may be taken as the estimate, T, of θ in our discussic above. We can then show that

$$p(z_1, z_2)dz_1dz_2 = 2dz_1dz_2, \tag{14}$$

$$p(z_2)dz_2 = 2\{1 - 2|z_2|\}dz_2 \cdots -\tfrac{1}{2} < z_2 < \tfrac{1}{2}, \tag{15}$$

and

$$p(z_1|z_2)dz_1 = \frac{dz_1}{1-2|z_2|} \cdots \theta - \tfrac{1}{2} + |z_2| < z_1 < \theta + \tfrac{1}{2} - |z_2|. \tag{16}$$

That these are the correct limits for z_1 and z_2 may be seen by reference to Fig, noting that z_1 and z_2 are constant along lines parallel to the respective diagona BD and AC of the square.

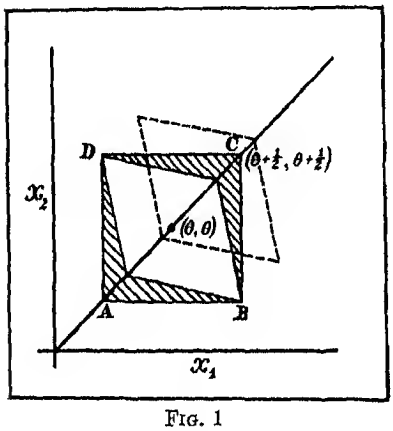


FIG. 1

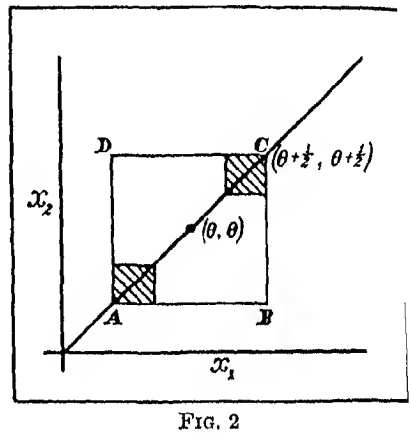


FIG. 2

First let us confine ourselves to samples with the same configuration z_2. Then, from (16), we can say that

$$P\{\theta - \alpha(\tfrac{1}{2} - |z_2|) < z_1 < \theta + \alpha(\tfrac{1}{2} - |z_2|)\} = \alpha. \tag{17}$$

This statement is true for given z_2, and will be *a fortiori* true when this restriction is removed. It is equivalent to saying that the chance of a point falling into the shaded area in Fig. 1 is $(1 - \alpha)$, where α denotes the proportion of the diagonal AC lying in the non-shaded area.[8] Confidence limits for θ are then obtained by transposing (17), giving

$$P\{z_1 - \alpha(\tfrac{1}{2} - |z_2|) < \theta < z_1 + \alpha(\tfrac{1}{2} - |z_2|)\} = \alpha. \tag{18}$$

[8] We are assuming that confidence limits are required such that the chance is $\left(\frac{1}{2} - \frac{\alpha}{2}\right)$ of θ being above the upper limit, and $\left(\frac{1}{2} - \frac{\alpha}{2}\right)$ of it being below the lower limit.

That this is not the best way of constructing confidence limits is seen as follows. Let us denote the lesser of x_1 and x_2 by x_L, and the greater by x_G. Then if we consider the possible values of x_L and x_G which will satisfy simultaneously the inequalities

$$(19) \qquad \begin{cases} \theta - \tfrac{1}{2} < x_L < \theta + \tfrac{1}{2} - \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \\ \theta - \tfrac{1}{2} + \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} < x_G < \theta + \tfrac{1}{2} \end{cases}$$

we see that they lie in the non-shaded area of the square $ABCD$ in Fig. 2 where the sides of the shaded squares are $\sqrt{\frac{1}{2} - \frac{\alpha}{2}}$. The chance of the inequalities holding simultaneously is therefore α. Further we see that these inequalities can be transposed to read

$$(20) \qquad \begin{cases} x_G - \tfrac{1}{2} < \theta < x_L + \tfrac{1}{2} \quad \text{when } (x_G - x_L) > \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \\ x_L - \tfrac{1}{2} + \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} < \theta < x_G + \tfrac{1}{2} - \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \\ \qquad\qquad \text{when } (x_G - x_L) < \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \end{cases}$$

and therefore we can take these to define our confidence limits for θ.

The intervals defined by the confidence limits in (18) and (20) are equivalent in the sense that each covers the true value of θ in a proportion α of cases. To decide which is the better rule of interval estimation we shall follow Neyman, and consider how often the intervals cover values other than the true θ. In particular let $(\theta + \Delta)$ be any other value, and consider the expressions P_1 and P_2 where

$$(21) \qquad P_1 = P\{z_1 - \alpha(\tfrac{1}{2} - |z_2|) < (\theta + \Delta) < z_1 + \alpha(\tfrac{1}{2} - |z_2|)$$

and P_2 is the probability that one or another of the following inequalities holds

$$(22) \qquad \begin{cases} x_G - \tfrac{1}{2} < (\theta + \Delta) < x_L + \tfrac{1}{2} \quad \text{when } (x_G - x_L) > \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \\ x_L - \tfrac{1}{2} + \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} < (\theta + \Delta) < x_G + \tfrac{1}{2} - \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \\ \qquad\qquad \text{when } (x_G - x_L) < \sqrt{\tfrac{1}{2} - \tfrac{\alpha}{2}} \end{cases}$$

Now (21) can be written

$$(23) \qquad P_1 = P\{(\theta + \Delta) - \alpha(\tfrac{1}{2} - |z_2|) < z_1 < (\theta + \Delta) + \alpha(\tfrac{1}{2} - |z_2|)\}.$$

Referring to Fig. 1 we see that we have to evaluate the chance of the sample falling into a lozenge-shaped area like the unshaded area in $ABCD$, but moved bodily along the diagonal AC to such a position as is indicated by the dotted lines. Difficulties are introduced by the discontinuities, but we can show that for Δ positive

$$(24)\quad \begin{cases} P_1 = \alpha \quad \text{when } \Delta = 0 \\ P_1 = \alpha - \dfrac{4\alpha\Delta^2}{1-\alpha^2} \cdots 0 \le \Delta \le \left(\dfrac{1}{2} - \dfrac{\alpha}{2}\right) \\ P_1 = \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) - 2\Delta + \dfrac{2\Delta^2}{1+\alpha} \cdots \left(\dfrac{1}{2} - \dfrac{\alpha}{2}\right) \le \Delta \le \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) \\ P_1 = 0 \qquad \Delta \ge \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) \end{cases}$$

with similar expressions for Δ negative. The graph of P_1 against Δ is shown in Fig. 3, α for convenience being taken = 0.92. From it we can read off the probability of the confidence interval covering $(\theta + \Delta)$, where θ is the true value of the parameter.

Similar calculations may be made for P_2. Without going into details, it is seen that

$$(25)\quad \begin{cases} P_2 = \alpha & \text{when } \Delta = 0 \\ P_2 = \alpha - 2\Delta\left(1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}}\right) & 0 \le \Delta \le \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ P_2 = \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) - 2\Delta + \Delta^2 & \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \le \Delta \le 1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ P_2 = 0 & \Delta \ge 1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \end{cases}$$

P_2 is plotted against Δ in Fig. 3. It is seen that, whatever value of Δ we take, the chance of $(\theta + \Delta)$ being included in the confidence interval, is less for the second method of estimation than it is for the method based on the distribution of $z_1 \mid z_2$.[9] This circumstance would, I think, contradict the view that the latter method was deriving the utmost from the sample. Whether the method is still a good one, though not necessarily the best, is not a question at issue in the present paper. The curves in Fig. 3 are very close together, and we are led to expect this by the fact that (12) is the weighted mean of the powers within the separate configurations, the weights being the probabilities $p(C)$ of the configurations. I am only concerned to show that certain methods, for which

[9] It will be noted that, when inverted, the curves of Fig. (iii) represent the power functions of tests for which the regions of rejection are those in figures (i) and (ii) respectively, the test being whether the parameter has the specified value θ, and different alternative hypotheses being represented by $(\theta + \Delta)$.

properties analogous to those of sufficiency have been claimed, do not satisfy conditions which I think they should, if these claims are to be upheld.

3. **Fiducial Distributions.** In the first section of this paper I discussed certain points of difference between the approaches to the problem of interval estimation made by R. A. Fisher on the one hand and J. Neyman and E. S. Pearson on the other The differences are not, perhaps, of the same magnitude as those between all these writers and the protagonists of inverse probability, and the results reached are so often the same that the reader may be excused for being somewhat impatient with what appear to be rather fine distinctions. However, as was seen in the last section, the approaches do not always yield exactly the

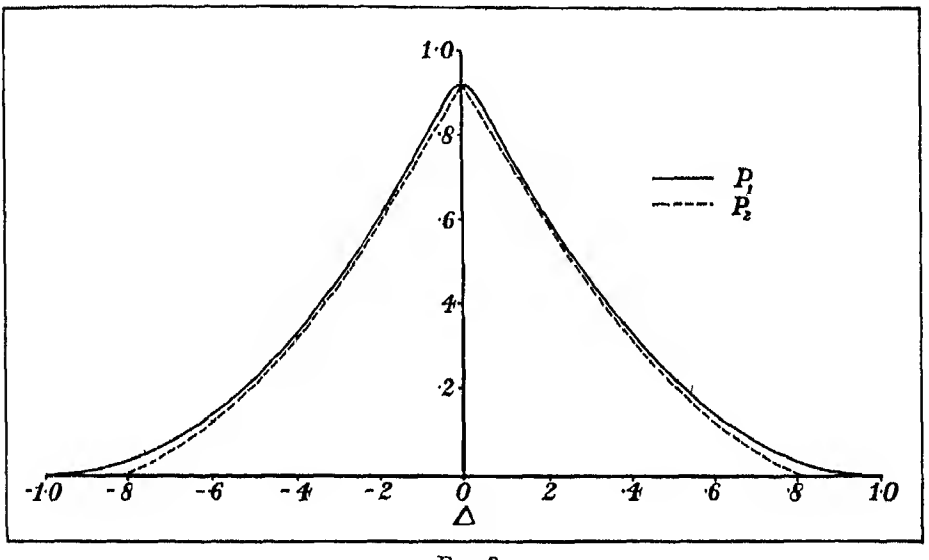


FIG. 3

same final results, and therefore I think it may be profitable to discuss them still further.

Closely connected with Fisher's desire to restrict the use of the fiducial method to situations where statistics exist which possess some property of sufficiency, is his introduction of the concept of a *fiducial distribution* for the unknown parameter. One can talk about *the* fiducial distribution for a parameter only if it is a unique distribution. Neyman, however, never makes use of fiducial distributions, and would, I think, claim that any valid results reached with the concept can equally well be reached without it. Where the results are the same there is room for two opinions on this matter. Some writers find it convenient to think in terms of fiducial distributions, and others prefer always to carry forward their reasoning as far as possible in terms of direct probability statements about the observational values, before transposing them to obtain confidence or fiducial limits for the parameters.

Greater objection can be made to the use of simultaneous fiducial distributions of several parameters. For instance, in the case of the normal distribution with parameters μ and σ, a simultaneous fiducial distribution has been defined in the following way.[10] Starting with the fact that the joint distribution of

$$\phi_1 = \frac{\sqrt{n}\,(\bar{x} - \mu)}{\sigma} \qquad \text{and } \phi_2 = \frac{(n-1)s^2}{\sigma^2}$$

is

$$df = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2})\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}\phi_1^2} e^{-\frac{1}{2}\phi_2} \phi_2^{\frac{1}{2}(n-3)}\, d\phi_1\, d\phi_2,$$

$\bar{x}$ and s are treated *formally* as fixed, and ϕ_1 and ϕ_2 are transformed to μ and σ, treated *formally* as variables. This gives

$$(26) \qquad df = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2})\Gamma\left(\frac{n-1}{2}\right)} \frac{\sqrt{n}}{\sigma} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} \cdot \frac{2}{\sigma} e^{-\frac{(n-1)s^2}{2\sigma^2}} \left\{\frac{(n-1)s^2}{\sigma^2}\right\}^{\frac{1}{2}(n-1)} d\mu\, d\sigma$$

This distribution would be useful if it were legitimate to integrate it out to obtain a fiducial distribution for any function $g(\mu, \sigma)$ say, of μ and σ. However, as for instance Bartlett has pointed out, this is not necessarily permissible. It seems to me therefore, that distributions defined as in (26) should be dispensed with entirely, for their very form encourages the belief that they can be integrated out at will. That this belief is still held is illustrated by a recent paper by Miss D. M. Starkey[11] concerned with the difference between the means of normal populations where the standard deviations are not assumed equal. This is the original problem to which Fisher[12] applied a method equivalent to integrating out the joint fiducial distribution of the two population means. Bartlett[13] raised an objection to this method of treatment, and I have also discussed the matter further.[14] Miss Starkey proceeds from the assumption that Fisher's method is sound.

The concept of the fiducial distribution has also been used in those problems of location and scaling, which have been treated by the procedure discussed above, of considering distributions in samples with the same configuration. Indeed it is one of the attractions of this procedure that we are led to distribu-

[10] R. A. Fisher, (1935). "The fiducial argument in statistical inference." *Ann Eugen* VI, p. 395.

[11] Daisy M. Starkey (1937). "A test of the significance of the difference between means of samples from two normal, populations without assuming equal variances." *Ann. Math. Stat.* Vol IX. No. 3, pp. 201–213.

[12] R. A. Fisher (1935). *loc. cit.*

[13] M S Bartlett (1936). "The information available in small samples." *Proc. Camb. Phil. Soc.* 32, pp. 560–566.

[14] B. L. Welch (1937). "The significance of the difference between two means when the population variances are unequal." *Biometrika*, XXIX, p. 358.

tions with, so to speak, one degree of freedom, so that the fiducial method may be safely applied. However, although probability statements based on such a fiducial method are here quite valid, I do not think that such statements can claim a *unique* validity. As I have shown in the previous section, there is no necessity to confine oneself to sampling within a configuration in order to obtain interval estimates for parameters, and we may fare better by not so confining ourselves, even if we have to dispense with the fiducial distribution.

4. **Summary.** Certain points which arise in the problem of estimating an interval in which a population parameter should lie have been discussed. In the second section it has been shown that in estimating location parameters it is not sufficient to consider the distribution of estimates in samples of the same configuration, meaning by sufficient that the sample is thereby utilized in the best possible way.

UNIVERSITY COLLEGE,
LONDON

THE REGRESSION SYSTEMS OF TWO SUMS HAVING RANDOM ELEMENTS IN COMMON

By J. F. Kenney

1. **Introduction.** The purpose of this note is to illustrate the power and elegance of the technique of characteristic functions[1] in solving a problem which has been discussed in the literature by Fischer[2] and others.

Let $x_1, x_2, \cdots, x_n$ be n variables independent of each other in the statistical sense, all subject to the same distribution function f, so that the function representing their joint distribution is

$$f(x_1)f(x_2)\cdots f(x_n). \tag{1}$$

Under these conditions a set of values $x_1, x_2, \cdots, x_n$ will be said to constitute a *sample of* n from a population with distribution function $f(x)$ and the function (1) will be said to represent the distribution of samples. It will be understood that $f(x)$ is defined and is non-negative for all real values of x and

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

If the actual occurrence of the variable is limited to a finite range, $f(x)$ is defined as identically zero outside that range.

The mathematical expectation of an arbitrary function $\psi(x)$, denoted by application of the operator E, is

$$E[\psi(x)] = \int_{-\infty}^{\infty} \psi(x)f(x)\,dx. \tag{2}$$

This integral will be convergent whenever $\psi(x)$ is absolutely integrable and bounded. In particular, if $\psi(x) = x$ we have the mean

$$a = \int_{-\infty}^{\infty} xf(x)\,dx$$

and it will be assumed that a exists.

Suppose a sample of n is taken from the population represented by $f(x)$ and the sum

$$y = x_1 + x_2 + \cdots + x_k + x_{k+1} + \cdots + x_n \tag{3}$$

---

[1] The writer takes pleasure in acknowledging his indebtedness to Professor A. T. Craig for suggesting this method.

[2] "On correlation surfaces of sums with a certain number of random elements in common," these Annals, vol. 4, no. 2, pp. 103–126.

is formed. From this sample $k < n$ values are chosen at random, and a sample of $m - k$ $(m \leq n)$ additional values, x_j', is taken from $f(x)$. The sum

$$z = x_1 + x_2 + \cdots + x_k + x_{k+1}' + \cdots + x_m' \tag{4}$$

is then formed. The problem is to determine the regression systems of z on y and y on z in the population resulting from repeated samples.

Before proceeding with the solution a brief discussion of characteristic functions will be given.

2. **Characteristic functions.** When $\psi(x) = e^{itx}$, where t is a real variable and $i = \sqrt{-1}$, (2) is called the characteristic function of x. Thus if we let $\varphi(t) = E(e^{itx})$ we have

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx.$$

From the conditions imposed on $f(x)$ it follows that the integral defining $\varphi(t)$ is convergent and $|\varphi(t)| \leq 1$. If the kth derivative of $\varphi(t)$ with respect to t exists we have

$$\left.\frac{d^k \varphi(t)}{dt^k}\right|_{t=0} = i^k \nu_k$$

where

$$\nu_k = \int_{-\infty}^{\infty} x^k f(x)\, dx.$$

Thus the characteristic function of x has the property that its kth derivative at the origin (divided by i^k) gives the kth moment of the distribution of x about the origin of x.

The notion of characteristic function extends readily to a distribution of several variables. In particular, let $F(y, z)$ be the joint distribution function of variables y and z subject to the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(y, z)\, dy\, dz = 1.$$

Then the characteristic function of $F(y, z)$ is

$$\varphi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it_1 y + it_2 z} F(y, z)\, dy\, dz \tag{5}$$

where y and z are defined in (3) and (4).

3. **Solution of the problem.** The distribution function associated with the population of samples is of the form given by (1). Consequently, the characteristic function of $F(y, z)$ can be written in the form

$$\int \cdots \int \prod_{j=1}^{k} e^{i(t_1+t_2)x_j} f(x_j)\, dx_j \prod_{j=k+1}^{n} e^{it_1 x_j} f(x_j)\, dx_j \prod_{j=k+1}^{m} e^{it_2 x_j'} f(x_j')\, dx_j'$$

the limits of integration being taken over all admissible values of the variables. The above expression reduces to

$$\varphi(t_1, t_2) = [\varphi(t_1 + t_2)]^{\frac{m}{k}}[\varphi(t_1)]^{n-k}[\varphi(t_2)]^{m-k}. \tag{6}$$

By the Fourier transform we have from (5),

$$F(y, z) = (1/2\pi)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it_1y-it_2z}\varphi(t_1, t_2)\, dt_1\, dt_2.$$

Since a distribution is completely determined by its characteristic function, $F(y, z)$ can be exhibited if $f(x)$ is known. However, the solution of the problem does not depend upon exhibiting $F(y, z)$.

Let $g(y)$ and $h(z)$ be the marginal distributions of y and z, respectively. Then the mean value of z for a fixed y is

$$\bar{z}_y = \int \frac{zF(y, z)}{g(y)}\, dz, \tag{7}$$

and the mean value of y for a fixed z is

$$\bar{y}_z = \int \frac{yF(y, z)}{h(z)}\, dy \tag{8}$$

where here and subsequently the integration is taken over all admissible values of the variables.

Let us now take the partial derivative of $\varphi(t_1, t_2)$, as given in (5), with respect to t_2 and evaluate the result at $t_2 = 0$. We obtain

$$\frac{\partial}{\partial t_2}\varphi(t_1, t_2)\Big|_{t_2=0} = \int\int ize^{it_1y}F(y, z)\, dy\, dz. \tag{9}$$

If we denote the left member of (9) by $G(t_1)$ and utilize (7) in the right member, (9) becomes

$$G(t_1) = \int g(y)\bar{z}_y ie^{it_1y}\, dy.$$

Application of the Fourier transform yields

$$ig(y)\bar{z}_y = \frac{1}{2\pi}\int e^{-it_1y}G(t_1)\, dt_1. \tag{10}$$

Now from (6),

$$G(t_1) = k\overline{\varphi(t_1)}^{n-1}\varphi'(t_1) + \overline{\varphi(t_1)}^n ia(m - k).$$

Therefore (10) may be written as follows,

$$ig(y)\bar{z}_y = \frac{k}{2\pi}\int e^{-it_1y}\overline{\varphi(t_1)}^{n-1}\varphi'(t_1)\, dt_1 + \frac{ia(m - k)}{2\pi}\int e^{-it_1y}\overline{\varphi(t_1)}^n\, dt_1. \tag{11}$$

To evaluate these integrals, consider

$$\overline{\varphi(t_1)}^n = \int e^{it_1 y} g(y)\, dy. \tag{12}$$

Differentiating (12) with respect to t_1 we have

$$n\overline{\varphi(t_1)}^{n-1}\varphi'(t_1) = \int iye^{it_1 y} g(y)\, dy. \tag{13}$$

Again using the Fourier transform, we obtain

$$iyg(y) = \frac{nk}{2k\pi}\int e^{-it_1 y}\overline{\varphi(t_1)}^{n-1}\varphi'(t_1)\, dt_1$$

from (13) and

$$g(y) = \frac{1}{2\pi}\int e^{-it_1 y}\overline{\varphi(t_1)}^{n}\, dt_1$$

from (12). Therefore (11) reduces to

$$iyg(y)\bar{z}_y = \frac{k}{n} iyg(y) + ia(m-k)g(y)$$

and we have at once the simple result

$$\bar{z}_y = ky/n + a(m-k). \tag{14}$$

In an analogous manner, it may be shown that

$$\bar{y}_z = kz/m + a(n-k). \tag{15}$$

Writing (14) and (15) in the forms

$$\begin{cases} \bar{z}_y - am = c_1(y - an) \\ \bar{y}_z - an = c_2(z - am) \end{cases} \tag{16}$$

where $c_1 = k/n$ and $c_2 = k/m$ are the regression coefficients it follows from the linearity of the regressions that the correlation coefficient is

$$\rho = \sqrt{c_1 c_2} = k/\sqrt{mn}.$$

If $m = n$, we have a well known result which is sometimes stated as follows: If y and z are affected by n equally likely causes of which k are common to both, then the correlation coefficient between y and z is equal to k/n.

NORTHWESTERN UNIVERSITY.

A NOTE ON CONFIDENCE INTERVALS AND INVERSE PROBABILITY

By Albert Wertheimer

The object of this note is to discuss a certain property of confidence intervals from the point of view of inverse probability. We shall not go into detailed applications, but merely into fundamental ideas, so we shall work with distribution functions that are continuous and satisfy conditions which are sufficient to insure the validity of the mathematical steps used.

A clear and concise statement of the subject is given in a paper by Neyman,[1] and we shall use it as the basis for our discussion. His presentation can be summarized as follows: Let x be a sample statistic having a distribution function

$$p(x, \theta) \qquad \begin{matrix} x_1 \leq x \leq x_2 \\ \theta_1 \leq \theta \leq \theta_2 \end{matrix}$$

where θ is a parameter of the population. Now define two monotonic functions

$$x = f(\theta); \qquad x = g(\theta) \qquad \begin{matrix} x_1 \leq x \leq x_2 \\ \theta_1 \leq \theta \leq \theta_2 \end{matrix}$$

such that $f(\theta) < g(\theta)$, and

$$\int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon, \qquad \text{for all } \theta. \tag{1}$$

Let the prior distribution function of θ be

$$\psi(\theta) \qquad \theta_1 \leq \theta \leq \theta_2 .$$

It then follows directly that the probability for any pair of values (x, θ) lying within the region enclosed by the curves is given by

$$\int_{\theta_1}^{\theta_2} \psi(\theta)\, d\theta \int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon. \tag{2}$$

regardless of the prior function $\psi(\theta)$. His conclusion then is this: Stating that

$$g^{-1}(x) \leq \theta \leq f^{-1}(x) \tag{3}$$

every time the observation gives us a value of x equal to that given in (3) we may in any one instance be wrong; this will happen only if the pair (x, θ) for this observation lies outside the region enclosed by the curves; but from (2) the probability for this to happen is ϵ. This statement is equivalent to saying that

[1] *Journal of the Royal Statistical Society*, Vol. 97, part IV, 1934; pp. 589–93.

if for every observed x we write the inequality (3), then for a large number of samples, the fraction $1 - \epsilon$ of the inequalities will be found correct.

We note here that this is true only if in the inequality (3) x is presumed to range over its entire interval of definition. But if for an observation $x = x'$, we mean to consider the corresponding inequality

$$g^{-1}(x') \leq \theta \leq f^{-1}(x') \tag{4}$$

as one member of the class of inequalities that could be written just for those samples that had $x = x'$, then we can not assert that the inequality (4) has a probability of $1 - \epsilon$ of being correct. In fact, any probability statement dealing with this class must involve the prior distribution function $\psi(\theta)$; and if it is not given, then we do not know in what percent of cases the restricted inequality (4) will be found correct.

Let us nevertheless approach the problem from the viewpoint of inverse probability. Having observed $x = x'$, the posterior probability of inequality (4) being correct is

$$\eta(x') = \frac{\int_{g^{-1}(x')}^{f^{-1}(x')} \psi(\theta)p(x', \theta)\, d\theta}{\int_{\theta_1}^{\theta_2} \psi(\theta)p(x', \theta)\, d\theta} \tag{5}$$

the numerator being the probability for the simultaneous occurrence of

$$x = x'; \qquad g^{-1}(x') \leq \theta \leq f^{-1}(x'),$$

and the denominator the probability[2] that $x = x'$, θ lying anywhere between θ_1 and θ_2.

As long as $\psi(\theta)$ is unknown $\eta(x')$ cannot be evaluated; however its average value $\bar{\eta}(x)$ with respect to x can be evaluated. By definition of an average,

$$\eta(x) = \int_{x_1}^{x_2} \eta(x)\, dx \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta \tag{6}$$

From (5) we have

$$\int_{g^{-1}(x)}^{f^{-1}(x)} \psi(\theta)p(x, \theta)\, d\theta = \eta(x) \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta \tag{7}$$

Integrating both sides of (7) over the entire range of x we get

$$\int_{x_1}^{x_2} dx \int_{g^{-1}(x)}^{f^{-1}(x)} \psi(\theta)p(x, \theta)\, d\theta = \int_{x_1}^{x_2} \eta(x)\, dx \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta$$

$$= \eta(x)$$

[2] When we say probability that $x = x'$, we mean the probability that x will lie in the internal $x \pm \frac{1}{2}dx$ to within terms of order dx.

Interchanging the order of integration, as is permissible under the assumptions, we get

$$\eta(x) = \int_{\theta_1}^{\theta_2} \psi(\theta)\, d\theta \int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx$$

But since

$$\int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon, \qquad \text{for all } \theta$$

we finally get

$$\bar{\eta}(x) = 1 - \epsilon$$

Thus when approached from the standpoint of inverse probability we see that the average value of the posterior probability of the inequality (4) is precisely the quantity $1 - \epsilon$ regardless of the prior distribution function $\psi(\theta)$.

In conclusion it is a pleasure to thank Dr. Deming for the criticisms and suggestions which he has made in connection with this note.

NAVY DEPARTMENT,
WASHINGTON.

NOTE ON A MATCHING PROBLEM

By Solomon Kullback

1. **Introduction.** There is to be found in the literature [1] a number of discussions of the matching problem i.e., the problem of deriving the distribution of the number of correct matchings when two sequences of elements are placed in correspondence. However, the formulation of the matching problem discussed and illustrated herein is somewhat different from those problems already discussed in the literature [1], and may be of interest. A rather general statement of the problem follows.

2. **The Problem.** Consider urns U_i, $i = 1, 2, \cdots, n$ each of which contains some or all of the r different elements $E_1, E_2, \cdots, E_r$. The relative proportions of the r elements in the i-th urn are $p_{i1}, p_{i2}, \cdots, p_{ir}$ $(i = 1, 2, \cdots, n)$ such that

$$p_{i1} + p_{i2} + \cdots + p_{ir} = 1 \qquad i = 1, 2, \cdots, n \tag{1}$$

$$p_{i1}^2 + p_{i2}^2 + \cdots + p_{ir}^2 = p_i \qquad i = 1, 2, \cdots, n \tag{2}$$

(Some p_i, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, r$ may be zero).

Assuming each urn to be an infinite source, consider two sequences made by drawing, at random, a single element from each urn in turn. If the two sequences are placed in correspondence there will be a number of correct matchings. What is the distribution of the number of correct matchings if the foregoing process be indefinitely repeated?

3. **Solution of the Problem.** The probability that the elements in the k-th position of the two sequences match may be derived by the following simple considerations Since all the drawings are independent, the probability that both elements in the k-th position are E_m is p_{km}^2. Accordingly, the probability that both elements are the same, irrespective of their particular identity is $p_{k1}^2 + p_{k2}^2 + \cdots + p_{kr}^2 = p_k$.

The theory for the number of correct matchings in this case thus corresponds to that for the Poisson series, which is well known [2]. For the special case in which $p_k = p$, $k = 1, 2, \cdots, n$ the distribution of the number of correct matchings is in accordance with the binomial $(q + p)^n$ where $q = 1 - p$.

4. **Numerical Illustration and Verification.** The following illustration corresponds to the special case in which the urns are taken to be identical with equal proportions of each of the r elements.

Random sequences of 300 digits each were matched and the number of correct matchings recorded. The result of 457 such observations is given in Table 1.

TABLE 1

Observed distribution of number of correct matchings per sequences of 300 random digits each

Number of correct matchings	Observed frequency	Number of correct matchings	Observed frequency
18	1	32	35
19	2	33	25
20	3	34	15
21	5	35	20
22	9	36	20
23	22	37	17
24	18	38	6
25	21	39	10
26	41	40	7
27	28	41	1
28	30	42	3
29	31	43	1
30	42	44	0
31	42	45	2
		Total	457

Average number of correct matchings	Standard deviation
29.9934	4.8484

TABLE 2

Values of $P_x = (300!/x!(300 - x)!)(0.1)^x(0.9)^{300-x}$

x	P_x	x	P_x	x	P_x	x	P_x
14	0.00033	23	0.03240	32	0.06920	41	0.00875
15	.00070	24	.04156	33	.06245	42	.00599
16	.00139	25	.05099	34	.05499	43	.00400
17	.00257	26	.05992	35	.04601	44	.00259
18	.00449	27	06756	36	.03763	45	.00164
19	.00741	28	07319	37	.02984	46	.00101
20	.01156	29	.07628	38	.02294	47	.00061
21	.01713	30	.07656	39	.01713	48	.00036
22	.02413	31	.07409	40	.01242	49	.00020

In accordance with paragraph 3, the distribution in Table 1 should correspond to the binomial distribution with $n = 300$ and $p = 10(1/10^2) = 1/10$. For the

TABLE 3

Comparison of observed distribution with the theoretical distribution $457\ (0.9 + 0.1)^{300}$

Number of correct matchings	Frequency		
	Observed	Theoretical	
	F_0	f	$F = 457f$
14–16	0	0.00242	1.1
17–19	3	.01447	6.6
20–22	17	05282	24.1
23–25	61	.12495	57.1
26–28	99	.20067	91 7
29–31	115	.22693	103.7
32–34	75	.18614	85.1
35–37	57	.11348	51.9
38–40	23	.05249	24.0
41–43	5	.01874	8 6
44–46	2	.00524	2.3
47–49	0	.00117	.5
	457		456.7

TABLE 4

F_0	F	$(F_0 - F)^2/F$	
0, 3	1.1, 6.6	2.87	
17	24.1	2.09	
61	57.1	.27	
99	91.7	.58	$\chi_0^2 = 10.48$
115	103.7	1.23	
75	85.1	1.20	
57	51.9	.50	$n = 8$
23	24.0	.04	
5, 2, 0	8.6, 2.3, .5	1.70	$P(\chi^2 > \chi_0^2) = .235$
		10.48	

binomial distribution we have $m = np = 30$, $\sigma = \sqrt{npq} = \sqrt{27} = 5.1962$. To compare the observed distribution with the expected distribution we calcu-

lated the values of $P_x = (300!/x!(300 - x)!)(0.1)^x(0.9)^{300-x}$ for values of x from 14 to 49 inclusive which are given in Table 2.

To compare the observed and the theoretical distributions, and test the "Goodness of Fit," the distributions were grouped in classes of three. The results are shown in Tables 3 and 4.

5. **Conclusion.** The agreement between the observed distribution and the theoretical distribution derived on the basis of the argument in paragraph 3 is quite satisfactory.

We have shown herein, that if two sequences be matched under certain conditions, the distribution of the number of correct matchings will, in general, be that of a Poisson series and in special cases the binomial distribution. The theory was illustrated by an experiment which yielded results in satisfactory agreement with the theory.

REFERENCES

[1] An indication of, and some of the problems discussed, will be found in: E. G. Olds, "A moment generating function which is useful in solving certain matching problems," *Bull. Am Math. Soc.*, Vol. 44, June 1938, pp. 407–413.

[2] Some referenees to discussions of the theory, moments, and distribution connected with the Poisson series follow:

(a) H. L Rietz, *Mathematical Statistics*, Carus Mathematical Monograph No. 3, Open Court Pub. Co (1927), pp 148–152

(b) Charles Jordan, *Statistique Mathematique*, pp. 109–110.

(c) T Kameda, "Theory of generating functions and its application to the theory of probability," *Journal of the Faculty of Science*, Imperial Univ. Tokyo, Section I, Vol I, Part 1, (1925), Theorem XXIV, p 48

(d) P. R. Rider, "The third and fourth moments of the generalized Lexis Theory," *Metron*, Vol 12, No 1, (1934) p. 195

(e) S. Kullback, On the Bernoulli Distribution, *Bull. Am Math Soc.*, Vol. 41, (1935) p. 861.

The George Washington University.

REPORT OF THE ANNUAL MEETING OF THE INSTITUTE

The fourth annual meeting of the Institute of Mathematical Statistics was held in Detroit, Michigan, on December 27–29, 1938, in conjunction with the meetings of the American Statistical Association and the Econometric Society. The program for the meetings was arranged by Professors S. S. Wilks and B. H. Camp.

On Tuesday morning, December 27, the Institute held a session devoted to contributed papers with Professor B. H. Camp, president of the Institute in the chair. At that time the following papers were presented:

1. *Generalizations of the Laplace-Liapounoff Theorem.*
   W. G. Madow, Millbank Management Corporation, New York.
2. *The standard errors of the geometric and harmonic means.*
   Nilan Norris, Hunter College.
3. *Note on an integral equation in population analysis.*
   Alfred J. Lotka, Metropolitan Life Insurance Company, New York.
4. *Optimum fiducial regions for simultaneous estimation of several population parameters from large samples.*
   S. S. Wilks, Princeton University.
5. *A mathematical contribution to immigration assessment.*
   Churchill Eisenhart, University of Wisconsin.
6. *Contributions to the theory of statistical estimation.*
   A. Wald, Columbia University.
7. *On the hypotheses underlying the applications of statistical methods to routine laboratory analyses.*
   J. Neyman, University of California.
8. *Commodity transformations and matrices.*
   Harold Hotelling, Columbia University.
9. *Remarks on two methods of sample inspection.*
   E. G. Olds, Carnegie Institute of Technology.

Abstracts of these papers are given at the close of this report.

Immediately following the session just described, the Institute convened in business session. At that time President Camp announced that the newly elected officers for the year 1939 are: President, P. R. Rider, Washington University; Vice-Presidents, C. C. Craig, University of Michigan, and S. S. Wilks, Princeton University; Secretary-Treasurer, A. T. Craig, University of Iowa.

The annual luncheon of the Institute was held at one o'clock on the same day. At the luncheon, Dr. Walter A. Shewhart, of the Bell Telephone Laboratories addressed the Institute on "The Future of Statistics in Mass Produc-

tion." A summary of this address is included among the abstracts at the close of this report.

On Wednesday morning, December 28, the Institute and the Statistical Association held a joint session devoted to the teaching of Business Statistics. Professor T. H. Brown presided. The following papers constituted the program:

1. *The teaching of undergraduate students.*
   L. S. Kellogg, Ohio State University.
2. *The teaching of graduate students.*
   O. W. Blackett, University of Michigan.
3. *A bead-sampling machine for use in the class room.*
   Dickson H. Leavens, Cowles Commission for Research in Economics.

Discussion: Harry P. Hartkemeier, University of Missouri.
Richard L. Kozelka, University of Minnesota.

On the afternoon of the same day, the Biometric Section of the Statistical Association and the Institute presented the following program on Statistical Methods in Genetics Problems with Professor Lowell J. Reed as chairman:

1. *Tests of simple Mendelian inheritance in randomly collected data of one and two generations.*
   Laurence H. Snyder and Charles W. Cotterman, Ohio State University.
2. *Statistical studies of the familial aspects of cancer in humans.*
   Herbert L. Lombard, Masachusetts State Department of Public Health.
3. *Application of the method of likelihood ratios to the testing of hypotheses of simple Mendelian inheritance.*
   S. S. Wilks, Princeton University.
4. *The application of statistical techniques to egg production data for the formulation of a breeding program.*
   W. C. Thompson, New Jersey Agricultural Experiment Station.

The Program Committees of the Institute and the Statistical Association arranged a joint session on Representative Sampling for Thursday afternoon, December 29. At that time the following papers were presented, with Professor Harold Hotelling presiding:

1. *On the mathematics of the representative method.*
   Allen T. Craig, University of Iowa.
2. *Application of the theory of sampling to large scale surveys and censuses.*
   Frederick F. Stephan, American Statistical Association.
3. *Further remarks on the mathematical aspects of representative sampling.*
   J. Neyman, University of California.

Discussion: Samuel A. Stouffer, University of Chicago.
Churchill Eisenhart, University of Wisconsin.
P. J. Rulon, Harvard University.

The final session of the meetings was held on Thursday evening. This was a joint session with the Econometric Society and was devoted to Mathematical

Statistics in Economics. Professor Irving Fisher presided and the following papers were given:

1. *On the hypothesis of linearity of regression in economic research.*
   J. Neyman, University of California.
2. *The selection of variates for use in prediction.*
   Harold Hotelling, Columbia University.
3. *Decomposition of time series on the basis of non-correlation principle.*
   Wassily Leontief, Harvard University.

Discussion: William G. Madow, Millbank Management Corporation, New York
Gerhard Tintner, Iowa State College.

A. T. CRAIG, *Secretary.*

ABSTRACTS OF PAPERS

(Presented on December 27, 1938, at the Detroit meeting of the Institute)

Generalizations of the Laplace-Liapounoff Theorem. W. G. MADOW, Milbank Management Corporation, New York.

The Laplace-Liapounoff Theorem states conditions under which a linear function of chance variables has a normal limiting distribution.

In dealing with limiting distributions arising in the analysis of variance, regression analysis, etc., there occurred problems which required for their solution the derivation of the joint limiting distribution of several linear functions of chance variables and the joint limiting distribution of functions which were linear in one set of chance variables for fixed values of other sets of chance variables.

These problems were solved by a matrix formulation of the Laplace-Liapounoff Theorem and by the introduction of a function whose convergence to zero in probability provided a sufficient condition for the existence of normal limiting distributions

Various generalizations with a view towards applications in multi-variate statistical analyses are discussed. The theorems provide a rigorous and complete basis for the derivation of limiting distributions of quadratic and bilinear forms

The Standard Errors of the Geometric and Harmonic Means. NILAN NORRIS, Hunter College.

Although certain properties of the geometric and harmonic means have been investigated extensively, there seems to have been no derivation of expressions for their variances in cases where they are used as estimates of parameters of parent populations.

Application of the modern theory of estimation makes it possible to develop simple and useful formulae for the standard errors of these two averages for each of the respective general classes of cases in which they are most suitable.

As in other instances in which standard errors are used in tests of significance, fiducial or confidence limits may be employed to overcome certain limitations of the outmoded practice of relying solely on multiples of either probable or standard errors to determine whether or not a result exists merely because of sampling fluctuations

Note on an Integral Equation in Population Analysis. ALFRED J. LOTKA, Metropolitan Life Insurance Company, New York.

In a population in which immigration and emigration are negligible, the number $N(t)$ of the population at time t is connected with the annual births $B(t)$ and the probability $p(a)$ of surviving from birth to age a, by the obvious relation

$$N(t) = \int_0^\infty B(t-a)p(a)\,da \tag{1}$$

If $B(t)$ and $p(a)$ are given, $N(t)$ follows at once by direct integration. The inverse problem, given $N(t)$, to find $B(t)$, requires separate treatment The case that $N(t)$ is given or can be expressed as a sum of exponential functions has been discussed by the

author on a former occasion. In the present communication it is shown how the function $B(t)$ can be expressed as a series proceeding in ascending derivatives of $N(t)$.

A second solution is also offered in which

$$\frac{B(t)}{N(t)} = b(t),$$

the birth rate per head is obtained as a series, the first and dominating term of which is

$$b(t) = \frac{1}{\int_0^\infty e^{-r_t a} p(a)\, da} \tag{2}$$

where r_t is the rate of natural increase at time t. This development is of interest because it corresponds to the expression for b in a population with constant birth rate, death rate, and rate of natural increase; that is

$$b = \frac{1}{\int_0^\infty e^{-ra} p(a)\, da} \tag{3}$$

so that the new expression represents $b(t)$ as the corresponding value of b in a Malthusian population, plus a series of correcting terms.

Optimum Fiducial Regions for Simultaneous Estimation of Several Population Parameters from Large Samples.

S. S. WILKS, Princeton University.

If a population has a distribution law $f(x, \theta)$ where x is the variate and θ is a parameter, it is known (*Annals of Mathematical Statistics*, Vol. IX (1938) pp. 166–175) that under rather general conditions, confidence intervals, for a given confidence coefficient α, which are shortest on the average, can be obtained from large samples of n items by solving the equations

$$\frac{\frac{\partial L}{\partial \theta}}{\sqrt{n}\sqrt{E\left[\frac{\partial \log f}{\partial \theta}\right]^2}} = \pm d_\alpha \tag{1}$$

for θ, where d_α is the normal deviate given by $\frac{1}{\sqrt{2\pi}}\int_{-d_\alpha}^{+d_\alpha} e^{-\frac{1}{2}t^2}\, dt = \alpha$. L is the logarithm of the likelihood, i.e. $L = \sum_{i=1}^{n} \log f(x_i, \theta)$, where E denotes mean value with respect to the probability law $f(x, \theta)$.

The present paper is an extension of the foregoing results to the case of several parameters. It is shown under fairly general conditions that if the distribution law of x if a function $f(x, \theta_1, \cdots, \theta_h)$ depending on h parameters, then for a confidence coefficient α the fiducial region of the θ's which is smallest in size *on the average* is given by the region in the space of the θ's for which

$$\frac{1}{n}\sum_{i,j=1}^{h} a_{ij}\left(\frac{\partial L}{\partial \theta_i}\right)\left(\frac{\partial L}{\partial \theta_j}\right) \leq \chi_\alpha^2 \tag{2}$$

where χ_α^2 is such that $P(\chi^2 \leq \chi_\alpha^2) = \alpha$, where the probability is calculated from a χ^2 distribution with h degrees of freedom. The matrix $|| a_{ij} ||$ is the inverse of the matrix whose general element is

$$E\left[\frac{\partial \log f}{\partial \theta_i} \cdot \frac{\partial \log f}{\partial \theta_j}\right]$$

Similar results hold when f is a function of several random variables as well as the parameters $\theta_1, \theta_2, \cdots, \theta_h$.

A Mathematical Contribution to Immigration Assessment. CHURCHILL EISENHART, University of Wisconsin.

A certain problem is assessing the size of an immigration can be stated mathematically as follows: A sample of size N is drawn at random from a population in which the probability of A is p. Let "not-A" be denoted by B. Then the sample will contain a frequency, say x, of A and $N - x$ of B, x being a random variable. This sample is now mixed together with a very much larger sample in which the elements are all B's, and the B's belonging to the original sample lost sight of. The problem is to estimate N from the observed frequency of A, namely x, in the composite sample, p being assumed known. The maximum likelihood estimate of N is x/p. For large values of x, and a fortiori of N, confidence intervals for N take the form $N_1 \leq N \leq N_2$ where

$$N_1 = \frac{\{\sqrt{q^2t^2 + 4q(x - \tfrac{1}{2})} - qt\}^2}{4pq},$$

$$N_2 = \frac{\{\sqrt{q^2t^2 + 4q(x + \tfrac{1}{2})} + qt\}^2}{4pq},$$

$$q = 1 - p,$$

and the confidence coefficient is .95 if t is set equal to 1.96, and is .99 if t is set equal to 2.58. For small values of x the solution is more difficult but charts are being prepared from which the confidence intervals can be read off.

A Contribution to the Theory of Statistical Estimation. A. WALD, Columbia University.

Let us denote by $f(x, \theta)$ a probability density function, where θ is a parameter. Denote by Ω the set of all possible values of θ. The assumption that θ belongs to a subset ω of Ω is called a hypothesis. Let us consider a system S of subsets of Ω. Denote the hypothesis corresponding to an element ω of S by H_ω and the system of all hypotheses corresponding to the elements of S by H_S. Denote by E a sample point in the n-dimensional sample space drawn from a population with the probability density function $f(x, \theta)$, where the value of θ is unknown. We have to decide by means of the sample point E which hypothesis of the system H_S should be accepted. That is to say, for each hypothesis H_ω we have to choose a region of acceptance M_ω in the sample space. The hypothesis H_ω will be accepted if and only if the sample point E falls in the region M_ω. Denote by M_S the system of all regions M_ω. The statistical problem to be solved is the question of *how the system of regions M_S should be chosen?*

In order to answer this question, a non-negative weight function $w(\theta, \omega)$ is introduced, which is defined for all values θ and for all elements ω of S. The weight $w(\theta, \omega)$ expresses the loss caused by accepting H_ω if θ is true. The probability of accepting H_ω multiplied by the weight $w(\theta, \omega)$ is called the risk of accepting H_ω if θ is true Denote this risk by

$r(\theta, H_\omega, M_S)$ (the risk depends obviously also on the system of regions M_S). The total risk of accepting a false hypothesis if θ is true, is given by

$$r(\theta, M_S) = \sum_{\omega} r(\theta, H_\omega, M_S)$$

where the summation is to be taken over all elements ω of S which do not contain θ.

Denote by $r(M_S)$ the maximum of $r(\theta, M_S)$ with respect to θ. The system M_S of regions for which $r(M_S)$ becomes a minimum is called the "best" system of regions relative to the weight function $w(\theta, \omega)$. Some properties of the best system of regions have been studied and the problem of its calculation has been treated.

On the Hypotheses underlying the Applications of Statistical Methods to Routine Laboratory Analyses.

J. NEYMAN, University of California.

The problem considered is that of estimating the proportion, p, of certain designated elements of the population sampled, the estimate to be based on a random sample of n, drawn by some mechanical device, such as is sometimes used in industry and in laboratory work. Examples: (1) to estimate the proportion of defective manufactured products in mass production; (2) to estimate the proportion of seeds which are able to germinate in given conditions. One would expect that the sample proportion, say q, will be distributed in repeated samples according to the Binomial Law and that, consequently, in order to obtain the confidence limits for p, one should use the Clopper-Pearson graphs. However, the evidence obtained from special analyses on seed germination, made in the Seed Testing Station of Warsaw, Poland, shows that this assumption may not be true. The sampling there was carried out by means of a machine and involved a certain amount of mixing. As a result q was more stable than it was expected. It did not follow the Binomial Law at all, but a Normal one about p, with a standard deviation, σ, which could be well estimated from the sum of squares of deviations from respective means. For a considerable period of time (18 months) σ retained a constant value (a characteristic of the action of the sampling machine) which was rather smaller than $(n^{-1}q(1-q))^{\frac{1}{2}}$.

Consequently, to have a preassigned frequency of correct statements concerning p, it was necessary to calculate the confidence intervals according to the formulae of the Normal Theory

$$q - t\sigma < p < q + t\sigma$$

with an appropriate value of t. Probably similar situations are rather common.

Remarks on Two Methods of Sampling Inspection.

E. G. OLDS, Carnegie Institute of Technology.

When the instructions for inspecting lots of size m specify that samples of size n be taken and the lot be passed without detailed inspection if no defectives are found, then, on the average, the maximum number of defectives are passed when the number of defects per lot is $\frac{m+1}{n+1}$ or $\frac{m+1}{n+1} - 1$.

If the quality of a lot is to be checked by drawing pieces until a fixed number of defective pieces are found, it is important to know that the expected number n_i necessary to obtain i defects is $i\frac{m+1}{p+1}$, where there are p defects in the lot. If $\frac{n_i}{i(m+1)}$ is used as an estimate of $\frac{1}{p+1}$, it is convenient to observe that the variance of $\frac{n_i}{i(m+1)}$ is

$$\frac{1}{p+2}\left[\frac{1}{p+1} - \frac{1}{m+1}\right]\left[\frac{1}{i} - \frac{1}{p+1}\right].$$

Commodity Transformations and Matrices. HAROLD HOTELLING, Columbia University.

If we regard the prices and quantities of n commodities as vectors we may apply the theory of linear transformations in various ways having economic and statistical significance. An example is the mixing of grains to produce results conforming to new specifications, as in international trade. Another kind of example arises in problems of multivariate statistical analysis such as those treated in my paper on "Relations between Two Sets of Variates" (*Biometrika*, 1936), concerned with properties invariant under internal linear transformations of the variates of each set. Prices transform contragrediently to quantities in all cases. Hence, if the prices and quantities of the same set of commodities are the two sets of variates, the allowable transformations are restricted. Consequently there are invariants in this case additional to those discussed in the paper mentioned Another problem is the reduction of sets of linear demand functions to normal form and determination of invariants when transformations of prices must be contragredient to those of commodities. The question whether the demand functions are symmetrical is here of paramount importance, since symmetry is preserved by such transformations, and since there are known theoretical reasons to expect symmetry. For a non-singular set of linear symmetrical demand *or* supply functions there are no invariants under arbitrary sets of contragredient transformations; but for pairs of such sets of demand *and* supply functions there are invariants, namely the elementary divisors of the pair of matrices of coefficients. A set of demand or supply functions alone has invariants if its matrix B is not symmetrical. If B' denote the transverse or conjugate of B, the elementary divisors of $B + \lambda B'$ are such invariants.

The Future of Statistics in Mass Production.[1] WALTER A. SHEWHART, Bell Telephone Laboratories, New York.

Much has been written about the application of statistical theory and technique in studying, discovering, and measuring the effects of an existing system of unknown or chance causes Much remains to be written about the application of statistical theory and technique in finding out how to tinker with and modify an existing chance cause system until it behaves as we would have it do. In research, we use statistical theory in helping to predict the future effects of some existing cause system. The statistician knows that his predictions will be valid if certain assumptions about the cause system are justified. Perhaps the most important assumption of this type is that the particular effects of a chance cause system under study are random. In mass production, however, the statistician has learned by experience that chance cause systems producing random effects don't just happen even under what we customarily consider to be the best regulated laboratory conditions. If the industrial statistician chooses to ignore this fact and makes predictions as if he were dealing with random cause systems, he may expect many of his predictions to fall far short of the truth: what is more, he knows that this fact will be discovered and his work discredited because in a continuing mass production process predictions are sure to be checked. Hence the industrial statistician in mass production must start not where the research statistician leaves off but, as it were, before the research statistician begins: that is, he must start by developing techniques for determining when we are justified in assuming that the underlying cause system is random. We thus arrive at a good starting point from which to consider the future of statistics in mass production.

Experience in the control of quality has provided a practical technique for detecting

[1] Summary of an address delivered at a luncheon meeting of the Institute of Mathematical Statistics, Detroit, Michigan, December 27, 1938.

and eliminating assignable causes of variability in the production process until a state of statistical control is reached where predictions based upon the assumption of randomness are likely to prove valid. It has also been shown elsewhere that by elimination of assignable causes of variability, we may make the most efficient use of raw materials, maximize the assurance of maintaining standard quality of manufactured goods, minimize the cost of inspection, and minimize the cost of rejections. Hence we may conclude that the use of statistics in mass production can be made to pay good dividends: such use can be made to have a bright future. On what then does that future depend?

To answer this question, we must consider the following three fundamental steps in the process of mass production:

I The specification of the quality of the thing wanted.

II. The production of things designed to meet the specification

III. The inspection of the things produced to see if they meet the specification.

An outstanding characteristic of the first step, specification, is the necessity of setting up and living within what we term a tolerance range[2] for each specified quality characteristic. If a producer contracts to live within some specified range and in taking steps II and III fails to do so, he usually loses a lot of money. Hence he must know how to set tolerance limits that he can meet Moreover, if he is to be able to make the most efficient use of materials in many instances, he must close up as much as he feasibly can on the specified tolerance range.

Obviously, however, one can not specify a practically attainable tolerance range out of thin air: one must be limited by what it is possible to do under commercial conditions of production in step II and this in turn is revealed by inspection in step III. We must also take into account the fact previously noted that any manufacturing process to begin with is almost certain not to be in a state of statistical control. In fact, this state can only be approached through the application of certain statistical techniques that have been found useful in detecting the presence of assignable causes that can be found and removed. A point to be stressed is that the three steps—specification, production, and inspection—in mass production, *cannot be taken independently: instead, they must be coordinated.* It also may be shown that maximum effectiveness in the use of statistical theory can only be attained by coördinating the applications in each of the three steps. It is significant to note that in order to attain the most efficient use of materials and processes by minimizing the tolerance range and in order to minimize the cost of production, one must make effective use of the results obtained in the course of commercial production, particularly those in the third step, inspection. In fact, the three steps might be thought of as constituting a scientific experiment in which the objective is the attainment of the most efficient use of available materials in the production of manufactured goods.

Broadly speaking, the statistician of the future has before him the opportunity of helping to develop this fundamental type of experiment in many respects like the way he is successfully helping today in so many fields of research to design experimental procedures that make the most efficient use of human effort Certain differences, of course, exist. For example, as already noted, he must start by designing a statistical control technique for randomizing, as it were, the cause system through the elimination of assignable causes. Then he can use modern statistical techniques of research in much the same way described in the literature with reasonable assurance that resulting predictions will be found valid because he has first randomized his cause system. He must, however, go farther than indicated in the current literature of statistical research in that he must provide operationally verifiable meanings for statistical terms such as random variable, accuracy, pre-

[2] The tolerance range is not to be confused with the fiducial range of modern statistics. The distinction between the two is set forth at some length in a forthcoming publication, *Statistical Method from the Viewpoint of Quality Control*, to be published shortly by the Graduate School of the United States Department of Agriculture

cision, true value, probability, degree of rational belief, and the like.[3] This is particularly necessary in the steps of specification and inspection because the specification is often made the basis of a contractual agreement between producers and consumers.

There is a sense in which the statistician's problem in helping to develop the mass production process so as to make the most effective use of information yielded by the process is much more complicated than the design of experiment usually considered in the literature of statistics. Whereas the customary statistical theory of design of experiment in research is concerned with comparatively small-scale experiments carried out under controlled conditions of the laboratory by a few people, the corresponding development of the mass production process must be carried out under commercial conditions on a large scale involving large numbers of people. For example, the three steps in the mass production process are usually carried out either by different companies or by different departments of the same company. The steps may involve the coördinated effort of literally hundreds and even thousands of employees, including physicists, chemists, engineers, sales agents, purchasing agents, lawyers, and economists. Very few of these people have ever had any training in statistics or probability and yet many of them must be sold on the use of statistics if the statistician is to have an opportunity of making his full contribution to the social and economic effectiveness of the mass production process. This situation constitutes a problem not only for those now in industry but also for those responsible for training the industrial leaders of tomorrow so that they will have sufficient knowledge of statistics to help them recognize the potential contributions of statistical theory and technique.

In conclusion, then, we may say that in the future the statistician in mass production must do more than simply study, discover, and measure the effects of existing chance cause systems: he must devise means for modifying these cause systems in the best way to satisfy human wants. The statistician in mass production must not be satisfied with simply measuring demand for goods; he must help change that demand by showing, among other things, how to close up the tolerance range and improve the quality of goods. He must not be content with measuring production costs; he must help decrease production costs through the use of the techniques of statistical control.

The future contribution of statistics in mass production lies not so much in solving the problems usually put to the statistician today by those not statistically trained as in taking a hand in helping to coordinate the steps of specification, production, and inspection considered as a scientific experiment for making the most efficient use of human effort in the production of goods to satisfy human wants The long range contribution of statistics depends perhaps not so much upon getting a lot of highly trained statisticians into industry in the immediate future as it does in creating a statistically minded new generation of those physicists, chemists, engineers, and others who will in any way have a hand in developing and directing the mass production process of tomorrow.

[3] An initial step in this direction has been taken in my Washington lectures. Loc. cit.

CONSTITUTION
OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE I

Name and Purpose

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

Membership

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.
2. Voting members of the Institute shall be (a) the Fellows, and (b) all others who have been members for twenty-three months prior to the date of voting.

ARTICLE III

Officers, Board of Directors, Committee on Membership, and Committee on Publications

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer, elected for a term of one year by a majority ballot at the annual meeting of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers and the previous President.

3. The Institute shall have a Committee on Membership composed of three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

4. The Institute shall have a Committee on Publications composed of three Members or Fellows elected by the Board of Directors. The President shall designate a Vice-President as Ex Officio Chairman of this Committee.

ARTICLE IV

Meetings

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such

time as the Board of Directors may designate. Additional meetings may be called from time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute.

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the Board by the Secretary-Treasurer at least five days prior to the date set therefor Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3. The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting.

4. At a regularly convened meeting of the Board of Directors, three members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

ARTICLE V

Publications

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. Other publications may be originated by the Board of Directors as occasion arises.

ARTICLE VI

Expulsion or Suspension

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

ARTICLE VII

Amendments

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

BY-LAWS

ARTICLE I

Duties of the Officers, Board of Directors, Committee on Membership, and Committee on Publications

1. The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present, shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be submitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2. The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute. He shall send out calls for annual dues and acknowledge receipt of same; pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditor shall report to the President.

3. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute.

4. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership.

5. The Committee on Publications, under the general supervision of the Board of Directors, shall have charge of all matters connected with the publications of the Institute, and of all books, pamphlets, manuscripts and other literary or scientific material collected by the Institute. Once a year this Committee shall cause to be printed in the Official Journal the Constitution and By-Laws and a classified list of all the Members and Fellows of the Institute.

ARTICLE II

Dues

1. Members shall pay five dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members shall pay

five dollars annual dues. The annual dues of Fellows shall be fivo dollars. The annual dues of Sustaining Members shall be fifty dollars. Honorary Members shall be exempt from all dues.

2. Annual dues shall be payable on the first day of January of each year.

3. The annual dues of a Fellow or Member include a subscription to the Official Journal. The annual dues of a Sustaining Member include two subscriptions to the Official Journal.

4. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues may be six months in arrears, and to accompany such notice by a copy of this Article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent one to the Board of Directors, by whom the person's name may be stricken from the rolls and all privileges of membership withdrawn. Such person may, however, be re-instated by the Board of Directors upon payment of the arrears of dues.

ARTICLE III

Salaries

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee.

ARTICLE IV

Amendments

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors.

DIRECTORY OF INSTITUTE OF MATHEMATICAL STATISTICS

(As of January 1, 1939)

Please notify the Secretary, A. T. Craig, University of Iowa, Iowa City, Iowa, of any changes of addresses or errors in names, titles or addresses as printed.

ACERBONI, DR. ARGENTINO V., Larroque 232 Banfield, Buenos Aires, Republic of Argentina, South America.

ALTER, PROFESSOR DINSMORE, Director of Griffith Observatory, Los Angeles, California.

ARNOLD, PROFESSOR HERBERT E., Department of Mathematics, Wesleyan University, Middletown, Connecticut.

AROIAN, PROFESSOR LEO, 522 Hill Street, Ann Arbor, Michigan.

BACON, PROFESSOR H M., Box 1144, Stanford, California.

BAKER, DR. G. A., Experiment Station, College of Agriculture, University of California, Davis, California..

BARRAL-SOUTO, DR. JOSÉ, Córdoba 1459, Buenos Aires, Republic of Argentina, South America.

BEALL, MR. GEOFFREY, 416 Queen Street, Chatham, Ontario, Canada.

BENNETT, PROFESSOR A. A., Department of Mathematics, Brown University, Providence, Rhode Island.

BERKSON, DR JOSEPH, The Mayo Clinic, Rochester, Minnesota.

BERNSTEIN, PROFESSOR FELIX, Department of Biometry, New York University, New York, New York..

BROOKS, MR. A. G., Western Electric Company, Hawthorne Station, Chicago, Illinois.

BURGESS, DR. R. W., Western Electric Co., 195 Broadway, New York, New York.

BURR, PROFESSOR IRVING W , Antioch College, Yellow Springs, Ohio.

BUSHEY, PROFESSOR J. HOBART, Department of Mathematics, Hunter College, Park Avenue and 68th Street, New York, New York.

CAMP, PROFESSOR B. H., Department of Mathematics, Wesleyan University, Middletown, Connecticut..

CAMP, PROFESSOR C. C., Department of Mathematics, University of Nebraska, Lincoln, Nebraska.

CARVER, PROFESSOR H. C., Department of Mathematics, The University of Michigan, Ann Arbor, Michigan..

CHAPMAN, MR ROY A., 400 Union Building, New Orleans, Louisiana.

CLARK, PROFESSOR ANDREW G., 1012 Laporte Avenue, Ft. Collins, Colorado

COLEMAN, MR. E. P., Department of Mathematics, The Municipal University of Omaha, Omaha, Nebraska.

COX, DR. GERALD J., The Mellon Institute, Pittsburgh, Pennsylvania.

CRAIG, PROFESSOR A. T., Department of Mathematics, The University of Iowa, Iowa City, Iowa..

CRAIG, PROFESSOR C. C , 3020 Angell Hall, The University of Michigan, Ann Arbor, Michigan..

CRATHORNE, PROFESSOR A. R., Department of Mathematics, The University of Illinois, Urbana, Illinois..

CROWE, PROFESSOR S. E., 137 University Drive, East Lansing, Michigan.
CURTISS, DR. J. H., Belle Ayre Apartments, Ithaca, New York.
DELURY, DR. D. B , Department of Mathematics, The University of Toronto, Toronto 5, Canada.
DEMING, DR. W E., Fertilizer Investigations, Friendship Post Office, Washington, D C..
DODD, PROFESSOR E. L , Department of Mathematics, The University of Texas, Austin, Texas..
DODGE, MR. HAROLD F., Bell Telephone Lab., 463 West Street, New York, New York.
DOOB, DR J. L., Department of Mathematics, The University of Illinois, Urbana, Illinois..
DUNLAP, PROFESSOR JACK W., Catharine Strong Hall, University of Rochester, Rochester, New York.
DWYER, PROFESSOR PAUL S , 407 Camden Court, Ann Arbor, Michigan.
EDGETT, DR. G. L., Queen's University, Kingston, Ontario, Canada.
EISENHART, DR. CHURCHILL, Department of Mathematics, The University of Wisconsin, Madison, Wisconsin.
ELSTON, MR. JAMES S., The Travelers Insurance Company, Hartford, Connecticut.
ELVEBACK, MISS MARY L , Room 11, Library, University of Minnesota, Minneapolis, Minnesota
EVANS, PROFESSOR H. P., North Hall, The University of Wisconsin, Madison, Wisconsin.
FEINLER, REV. F. J., 821 East C Street, Grants Pass, Oregon.
FELDMAN, DR. H. M., Soldan High School, St. Louis, Missouri.
FERTIG, DR. JOHN W., Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Maryland.
FISCHER, DR. C. H., Department of Mathematics, Wayne University, Detroit, Michigan.
FISHER, PROFESSOR IRVING, 460 Prospect Street, New Haven, Connecticut..
FOSTER, MR. RONALD M , 122 East Dudley Avenue, Westfield, New Jersey
FRANKEL, MR. LESTER R., 1445 Otis Place NW, Washington, D. C.
FRY, DR. T. C., 463 West Street, New York, New York..
GAVETT, PROFESSOR G. I , Department of Mathematics, The University of Washington, Seattle, Washington.
GILL, MR. JOHN P., Department of Mathematics, The University of Alabama, University, Alabama.
GIRSHICK, MR. M A , Bureau of Home Economics, Department of Agriculture, Washington, D C.
GLOVER, PROFESSOR J. W , 620 Oxford Road, Ann Arbor, Michigan
GREENWOOD, DR. J. A., Box 4723, Duke University, Durham, North Carolina
GREVILLE, DR. T. N. E , Department of Mathematics, The University of Michigan, Ann Arbor, Michigan
GROVE, PROFESSOR C. C., 143 Melburn Avenue, Baldwin, L. I., New York.
HART, PROFESSOR W. L , Department of Mathematics, The University of Minnesota, Minneapolis, Minnesota.
HENDERSON, MR. ROBERT, Crown Point, Essex County, New York..
HENDRICKS, MR WALTER A., National Agricultural Research Center, Beltsville, Maryland.
HENRY, MR M. H., 829 West Allegan Avenue, Lansing, Michigan.
HOEL, DR. PAUL G., Department of Mathematics, Oregon State College, Corvallis, Oregon.
HOTELLING, PROFESSOR HAROLD, Fayerweather Hall, Columbia University, New York, New York..
HUNTINGTON, PROFESSOR E V., 48 Highland Street, Cambridge, Massachusetts..
INGRAHAM, PROFESSOR M. H., North Hall, The University of Wisconsin, Madison, Wisconsin .
JACKSON, PROFESSOR DUNHAM, 119 Folwell Hall, The University of Minnesota, Minneapolis, Minnesota..
JANKO, DR. JAROSLAV, Praha XII, Krkonošská 3, Czechoslovakia

JURAN, MR. J. M., Western Electric Co , 195 Broadway, New York, New York.
KEEPING, PROFESSOR E. S , University of Alberta, Edmonton, Alberta, Canada.
KELLEY, PROFESSOR TRUMAN L., Lawrence Hall, Cambridge, Massachusetts .
KENNEY, MR. JOHN F., Department of Mathematics, Northwestern University, Evanston, Illinois.
KIERNAN, PROFESSOR CHARLES J., 19 Looker Street, Hillside, New Jersey.
KNOWLER, DR. LLOYD A., Department of Mathematics, Hunter College, New York, New York.
KOSSACK, MR. CARL F., RFD #2, Ann Arbor, Michigan.
KULLBACK, DR SOLOMON, Office of Chief Signal Officer, Washington, D. C..
KURTZ, DR. ALBERT K , 173 Cornwall Street, Hartford, Connecticut
LARSEN, PROFESSOR HAROLD D., Department of Mathematics, University of New Mexico, Albuquerque, New Mexico.
LEAVENS, MR. DICKSON H., 301 Mining Exchange Building, Colorado Springs, Colorado.
LEMME, PROFESSOR MAURICE M , Department of Mathematics, University of Toledo, Toledo, Ohio.
LIVERS, MR. JOE J., 723 Oakland Avenue, Ann Arbor, Michigan.
LORGE, DR. IRVING, 525 West 120 Street, New York, New York.
LOTKA, DR A. J , One Madison Avenue, New York, New York.
MADOW, DR. WILLIAM G., Millbank Management Corporation, Chrysler Building, New York, New York.
MALZBERG, DR. BENJAMIN, New York State Department of Mental Hygiene, Albany, New York.
MAUCHLY, PROFESSOR J. W., Department of Physics, Ursinus College, Collegeville, Pennsylvania.
MODE, PROFESSOR E. B., 688 Boylston Street, Boston, Massachusetts.
MOLINA, MR. E. C., 463 West Street, New York, New York..
MUDGETT, PROFESSOR BRUCE D., Department of Economics, University of Minnesota, Minneapolis, Minnesota.
MACLEAN, MR. M. C., Chief of Census Analysis, Dominion Bureau of Statistics, Ottawa, Canada.
MCEWEN, PROFESSOR G. F., Scripps Institution, La Jolla, California
NESBITT, DR. C. J., Department of Mathematics, The University of Michigan, Ann Arbor, Michigan.
NEYMAN, PROFESSOR J., Department of Mathematics, The University of California, Berkeley, California..
NORRIS, DR. NILAN, Department of Economics, Hunter College, New York, New York.
OLDS, PROFESSOR E G., 953 La Clair Avenue, Regent Square, Pittsburgh, Pennsylvania.
OLLIVIER, DR. ARTHUR, Box 405, State College, Mississippi.
OLPIN, MR. A R., Kendall Mills, Paw Creek, North Carolina.
O'TOOLE, DR. A. L., 1425 E. 60th Street, Chicago, Illinois.
ORE, PROFESSOR OYSTEIN, Department of Mathematics, Yale University, New Haven, Connecticut
PARENTE, MR. A. R., 126 Church Street, Hamden, Connecticut.
PIXLEY, PROFESSOR H. H., Department of Mathematics, Wayne University, Detroit, Michigan.
POLLARD, PROFESSOR H. S., Department of Mathematics, Miami University, Oxford, Ohio.
REED, PROFESSOR L. J., 615 North Wolfe Street, Baltimore, Maryland
RIDER, PROFESSOR P. R., Department of Mathematics, Washington University, St. Louis, Missouri..
RIETZ, PROFESSOR H. L., Department of Mathematics, The University of Iowa, Iowa City, Iowa..
ROMIG, MR. H. G., 463 West Street, New York, New York.

ROOS, DR. CHARLES, Care of Mercer-Allied Corporation, 420 Lexington Avenue, New York, New York..
RULON, PROFESSOR P. J., 4 Emerson Hall, Harvard University, Cambridge, Massachusetts.
SCARBOROUGH, PROFESSOR J. B., Post Office Box 332, Annapolis, Maryland.
SHEWHART, DR. W. A., 158 Lake Drive, Mountain Lakes, New Jersey..
SHOHAT, PROFESSOR J., Department of Mathematics, The University of Pennsylvania, Philadelphia, Pennsylvania..
SPIEGELMAN, MR. MORTIMER, 325 West 86th Street, New York, New York.
STEPHAN, MR. F. F., 1626 K Street NW, Washington, D. C.
STOUFFER, PROFESSOR S. A., Department of Sociology, University of Chicago, Chicago, Illinois.
SWANSON, MR A. G., General Motors Institute, Flint, Michigan.
THOMPSON, DR. WILLIAM R., 27 Oakwood Street, Albany, New York
TOOPS, PROFESSOR HERBERT A., The Ohio State University, Columbus, Ohio.
TORREY, MISS MARY N., 463 West Street, New York, New York.
TRELOAR, PROFESSOR ALAN E., Department of Biometry, The University of Minnesota, Minneapolis, Minnesota.
TRUBRIDGE, DR. G. F. P., 98 Kingsway, Petts Wood, Kent, England.
VATNSDAL, MR. J. R., 304 Columbia Street, Pullman, Washington.
VICKERY, DR. C. W., 304 E. Fifth Street, Austin, Texas.
WALKER, PROFESSOR HELEN M., Teachers College, Columbia University, New York, New York.
WEI, MR. DZUNG-SHU, 19 Evans Street, Iowa City, Iowa.
WEIDA, PROFESSOR F. M., Department of Mathematics, The George Washington University, Washington, D C..
WESCOTT, MR. MASON E., 1936 Greenwood Avenue, Wilmette, Illinois.
WELKER, MR. E. L., 160 Mathematics Building, University of Illinois, Urbana, Illinois.
WHITE, PROFESSOR A. E., Department of Mathematics, Kansas State College, Manhattan, Kansas.
WILKS, PROFESSOR S. S., Fine Hall, Princeton University, Princeton, New Jersey..
WILSON, DR. ELIZABETH W., One Waterman Street, Cambridge, Massachusetts.
WILSON, MR. W. P., Traffic Engineering Bureau, 206 Police Headquarters, City of Detroit, Detroit, Michigan.
WRIGHT, PROFESSOR SEWALL, Department of Zoology, University of Chicago, Chicago, Illinois..
WYCKOFF, MR. J. F., Department of Mathematics, Trinity College, Hartford, Connecticut.
ZOCH, MR. RICHMOND T., Cosmos Club, 1520 H Street, Washington, D. C.

ON THE SAMPLING THEORY OF ROOTS OF DETERMINANTAL EQUATIONS

By M. A. Girshick[1]

In a recent paper[2] Hotelling has considered two functions of the covariances of two sets of variates (having a multivariate normal distribution with s variates in the first set, t variates in the second, $s \leq t$) which he designates by Q and Z and which he defines as follows:

$$Q^2 = \frac{(-1)^s C}{AB} \quad \text{and}^3 \quad Z = \frac{D}{AB} \tag{1.1}$$

where A is the determinant of the covariances among the variates of the first set, B the determinant of the covariances among the variates of the second set, D the determinant of covariances of the two sets taken together, and C a determinant obtained from D by replacing the covariances among the variates of the first set by zeros. Both Q^2 and Z are shown to be invariant under internal linear transformations of either set of variates.

In solving the problem of determining linear functions of the two sets of variates for which the multiple correlation is a maximum, Hotelling arrives at a set of parameters $\rho_1, \rho_2, \cdots, \rho_s$ which he names "canonical correlations" and which are the positive or zero roots of the determinantal polynomial

$$D(\lambda) = \begin{vmatrix} -\lambda\sigma_{11} & \cdots & -\lambda\sigma_{1s} & \sigma_{1,s+1} & \cdots & \sigma_{1,s+t} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\lambda\sigma_{s1} & \cdots & -\lambda\sigma_{ss} & \sigma_{s,s+1} & \cdots & \sigma_{s,s+t} \\ \sigma_{s+1,1} & \cdots & \sigma_{s+1,s} & -\lambda\sigma_{s+1,s+1} & \cdots & -\lambda\sigma_{s+1,s+t} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{s+t,1} & \cdots & \sigma_{s+t,s} & -\lambda\sigma_{s+t,s+1} & \cdots & -\lambda\sigma_{s+t,s+t} \end{vmatrix}. \tag{1.2}$$

The ρ's are equal in number to the variates of the first set and bear the following relations to Q and Z:

$$Q^2 = \rho_1^2 \rho_2^2 \cdots \rho_s^2 \tag{1.3}$$

$$Z = (1 - \rho_1^2)(1 - \rho_2^2) \cdots (1 - \rho_s^2). \tag{1.4}$$

The corresponding functions for the sample covariances Hotelling designates by q and z, and the sample canonical correlations by $r_1, r_2, \cdots, r_s$. Under the assumption of complete independence between the two sets of variates and

---

[1] Most of this Research was accomplished at Columbia University under a Grant-in-Aid from the Carnegie Corporation of New York

[2] Harold Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, Vol XXVIII, Dec. 1936.

[3] The function Z was first considered by S. S. Wilks in *Biometrika*, Vol XXIV, Nov 1932

in the case $s = 2$ and $t = 2$, he shows that the joint distribution of q and z is of the form

$$\tfrac{1}{2}(n-2)(n-3)z^{\frac{1}{2}(n-5)}\,dq\,dz \tag{1.5}$$

q and z satisfying the inequalities

$$0 \leq z \leq 1, \qquad 0 \leq q \leq 1, \qquad z \leq (1-q)^2$$

and the joint distribution of the canonical correlations r_1 and r_2 is of the form

$$(n-2)(n-3)(r_1^2 - r_2^2)(1-r_1^2)^{\frac{1}{2}(n-5)}(1-r_2^2)^{\frac{1}{2}(n-5)}\,dr_1\,dr_2 \tag{1.6}$$

where n is one less than the number in the sample for each variate.

I

In Part I of this paper we shall, assuming independence between the two sets, find the joint moments of q and z for a general value of s and t and extend the joint distribution of q and z and hence of the canonical correlations to the case where there are two variates in the first set and any number of variates in the second, i.e. $s = 2$ and $t > 2$.[4]

1. **Joint Moments of q and z.** Since we are assuming complete independence between the two sets of variates we may without any loss of generality represent the sample values of the second set as points on the first t axes of unit distance from the origin in a space of n dimensions The matrix of observations in the case of s variates in the first set and t variates in the second set will take the form

$$\left\|\begin{array}{cccccccc} x_{11} & x_{12} & x_{13} & \cdots & x_{1t} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2t} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ x_{s1} & x_{s2} & x_{s3} & \cdots & x_{st} & \cdots & x_{sn} \\ 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \end{array}\right\| \tag{1.7}$$

The polynomial $D(\lambda)$ of (1.2) in terms of sample variances and covariances calculated from (1.7) then becomes

$$D(\lambda) = \left|\begin{array}{cccccc} -\lambda a_{11} & \cdots & -\lambda a_{1s} & x_{11} & \cdots & x_{1t} \\ \cdot & & \cdot & \cdot & & \cdot \\ -\lambda a_{s1} & \cdots & -\lambda a_{ss} & x_{s1} & \cdots & x_{st} \\ x_{11} & \cdots & x_{s1} & -\lambda & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ x_{1t} & \cdots & x_{st} & 0 & \cdots & -\lambda \end{array}\right| \tag{1.8}$$

where $a_{ij} = \sum_{1}^{n} x_i x_j$.

[4] This extension is a generalization of Hotelling's method loc. cit.

We multiply the first s rows of (1.8) by λ and factor out λ from the last t columns. This yields

$$D(\lambda) = \lambda^{t-s} \begin{vmatrix} -\lambda^2 a_{11} & \cdots & -\lambda^2 a_{1s} & x_{11} & \cdots & x_{1t} \\ \cdot & & \cdot & \cdot & & \cdot \\ -\lambda^2 a_{s1} & \cdots & -\lambda^2 a_{ss} & x_{s1} & \cdots & x_{st} \\ x_{11} & \cdots & x_{s1} & -1 & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ x_{1t} & \cdots & x_{st} & 0 & \cdots & -1 \end{vmatrix}. \tag{1.9}$$

As a further simplification, we multiply the $(s+j)^{\text{th}}$ column by $x_{i,}$ for all j from 1 to t and add the result to the i^{th} column. When this is done for every value of i from 1 to s and the resulting determinant expanded by means of the last t columns, the determinantal polynomial (1.9) becomes

$$D(\lambda) = \lambda^{t-s} \begin{vmatrix} b_{11} - \lambda^2 a_{11} & b_{12} - \lambda^2 a_{12} & \cdots & b_{1s} - \lambda^2 a_{1s} \\ \cdot & \cdot & \cdots & \cdot \\ b_{s1} - \lambda^2 a_{s1} & b_{s2} - \lambda^2 a_{s2} & \cdots & b_{ss} - \lambda^2 a_{ss} \end{vmatrix}$$

or symbolically

$$D(\lambda) = \lambda^{t-s} \, | \, b_{ij} - \lambda^2 a_{ij} \, | \tag{1.10}$$

where $b_{ij} = \sum_{1}^{t} x_i x_j$.

Hence the s roots of $D(\lambda)$ which do not necessarily vanish may be obtained from the polynomial

$$Q(\lambda) = | \, b_{ij} - \lambda^2 a_{ij} \, |. \tag{1.11}$$

The coefficient of the highest power of λ in $Q(\lambda)$ is given by $| \, a_{ij} \, |$, the determinant of the elements a_{ij}. Taking this in conjunction with (1.3) and (1.4) we see that

$$q^2 = \frac{Q(0)}{|\, a_{ij} \,|} = \frac{|\, b_{ij} \,|}{|\, a_{ij} \,|}$$
$$z = \frac{Q(1)}{|\, a_{ij} \,|} = \frac{|\, c_{ij} \,|}{|\, a_{ij} \,|} \tag{1.12}$$

where $c_{ij} = \sum_{i+1}^{n} x_i x_j$.

From the equations (1.12) we obtain

$$E\{ | \, a_{ij} \, |^{\frac{1}{2}(\alpha+2\beta)} q^{\alpha} z^{\beta} \} = E\{ | \, b_{ij} \, |^{\frac{1}{2}\alpha} | \, c_{ij} \, |^{\beta} \} \tag{1.13}$$

where E stands for the mathematical expectation of the expressions in the $\{\ \}$.

It is obvious from the definition of b_{ij} and c_{ij} that the two determinants $| \, b_{ij} \, |$ and $| \, c_{ij} \, |$ are independently distributed. Moreover, the joint distribution of q and z does not depend on the determinant $| \, a_{ij} \, |$. The truth of the latter statement can be seen from the following geometrical considerations. If we con-

sider the sample values of each variate as a point in an n-dimensional space, then the two sets of variates determine two flat spaces, one of s dimensions and one of t dimensions in that space A sample canonical correlation can then be considered as the cosine of a certain minimum or stationary angle between two lines, one line lying in the flat s space and the other in the flat t space. Since q and z are functions of the canonical correlations, they therefore depend only on lines and angles between two planes. The quantities a_{ij} on the other hand, depend on lines and angles lying entirely within one of these planes.

From the above considerations we see that equation (1.13) can be written as

$$E\{|\, a_{ij} \,|^{\frac{1}{2}(\alpha+2\beta)}\}E(q^\alpha z^\beta) = E(|\, b_{ij} \,|^{\frac{1}{2}\alpha})E(|\, c_{ij} \,|^\beta)$$

or

$$E(q^\alpha z^\beta) = \frac{E(|b_{ij}|^{\frac{1}{2}\alpha})E(|\, c_{ij}|^\beta)}{E(|\, a_{ij}|^{\frac{1}{2}(\alpha+2\beta)})} \tag{1.14}$$

The m^{th} moment of a determinant $|\, d_{ij} \,|$ of sums of sample cross products of p variates is given by the formula[5]

$$E(|\, d_{ij} \,|^m) = \frac{2^{pm}}{|D_{ij}|^m} \prod_{i=1}^{p} \left[\frac{\Gamma\left(\frac{n+2m+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)} \right], \tag{1.15}$$

where D_{ij} denotes the cofactor corresponding to σ_{ij} divided by the determinant $|\, \sigma_{ij} \,|$. Substituting (1.15) in (1.14) and simplifying, we get for the joint moments of q and z

$$E(q^\alpha z^\beta) = \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{t+\alpha+1-i}{2}\right)\Gamma\left(\frac{n-t+2\beta+1-i}{2}\right)\Gamma\left(\frac{n+1-i}{2}\right)}{\Gamma\left(\frac{t+1-i}{2}\right)\Gamma\left(\frac{n-t+1-i}{2}\right)\Gamma\left(\frac{n+\alpha+2\beta+1-i}{2}\right)} \right]. \tag{1.16}$$

2 Joint Distribution of q and z for $s = 2$, $t > 2$. In order to determine the joint distribution of q and z for $s = 2$ and $t > 2$, we shall first prove the following lemma.

LEMMA: *Let q and z be defined as in* (1.1) *for two sets of variates having s variates in either set and let q' and z' be similarly defined with $s < t$ where s is the number of variates in the first set and t the number of variates in the second set, then for $n = t + s$, the joint distribution of q^2 and z is identical with that of z' and q'^2.*

PROOF. If the number of variates in either set are the same and $n = t + s$, then by (1.12)

$$q^2 = \frac{|\, b_{ij} \,|}{|\, a_{ij} \,|}, \qquad z = \frac{|\, c_{ij} \,|}{|\, a_{ij} \,|}$$

[5] Cf. S. S Wilks, "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. XXIV, Nov. 1932.

where

$$b_{ij} = \sum_{1}^{t} x_i x_j, \qquad c_{ij} = \sum_{s+1}^{t+s} x_i x_j, \qquad a_{ij} = \sum_{1}^{t+s} x_i x_j \tag{1.17}$$

and $s = t$.

However, for $s < t$, and $n = t + s$, we take for the second set of t variates points on the t axes at unit distance from the origin in the $(t + s)$-dimensional space *perpendicular* to the first s axes. The matrix of observations in this case takes the form

$$\begin{Vmatrix} x_{11} & x_{12} \cdots x_{1s} & x_{1,s+1} \cdots x_{1,s+t} \\ x_{21} & x_{22} \cdots x_{2s} & x_{2,s+1} \cdots x_{2,s+t} \\ \cdot & \cdot \cdot \cdot & \cdot \cdots \cdot \\ x_{s1} & x_{s2} \cdots x_{ss} & x_{s,s+1} \cdots x_{s,s+t} \\ 0 & 0 \cdots 0 & 1 \cdots 0 \\ \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot \\ 0 & 0 \cdots 0 & 0 \cdots 1 \end{Vmatrix}. \tag{1.18}$$

Employing the same arguments as in equations (1.8) (1.9) and (1.10) we find that

$$Q(\lambda) = |c_{ij} - \lambda^2 a_{ij}|, \qquad q'^2 = \frac{|c_{ij}|}{|a_{ij}|}, \qquad z' = \frac{|b_{ij}|}{|a_{ij}|} \tag{1.19}$$

where

$$b_{ij} = \sum_{1}^{t} x_i x_j, \qquad c_{ij} = \sum_{s+1}^{t+s} x_i x_j, \qquad a_{ij} = \sum_{1}^{t+s} x_i x_j.$$

Comparing these equations with (1.17) we see that

$$z = q'^2, \qquad q^2 = z'. \tag{1.20}$$

This proves the lemma.

Now let $s = 2$. Setting $n = t + 2$ in equation (1.5) and using the transformation (1.20) we get for the joint distribution of q' and z'

$$\tfrac{1}{2}t(t-1)q'^{t-2}z'^{-\frac{1}{2}}\,dq'\,dz'. \tag{1.21}$$

Let r be the correlation between the two variates of the first set. The distribution of r in samples for which $n = t + 2$ when the population correlation is zero is known to be

$$\frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}}(1 - r^2)^{\frac{1}{2}(t-1)}\,dr. \tag{1.22}$$

The distribution of r is independent of q and z Hence, the joint distribution of q', z', and r is given by the product of (1.21) and (1.22). Dropping the

primes from q' and z' in (1.21), we get for the joint distribution of the three quantities in the case $n = t + 2$,

$$\frac{1}{2}t(t-1)\frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}}q^{t-2}z^{-\frac{1}{2}}(1-r^2)^{\frac{1}{2}(t-1)}\,dq\,dz\,dr. \tag{1.23}$$

We shall now derive the joint distribution of q and z for a general value of n for $s = 2$, $t > 2$. We set $x_1 = x$, $x_2 = y$ and take the t sample variates of the second set to be points on the first t axes at unit distance from the origin in a space of n dimensions. As in (1.12) calculate q and z.

$$q^2 = \frac{\sum_1^t x^2 \sum_1^t y^2 - \left(\sum_1^t xy\right)^2}{1-r^2}, \qquad z = \frac{\sum_{t+1}^n x^2 \sum_{t+1}^n y^2 - \left(\sum_{t+1}^n xy\right)^2}{1-r^2}. \tag{1.24}$$

We transform the points $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ to hyperspherical coordinates, the transformation to be represented parametrically by the equations

$$\begin{aligned}
x_1 &= \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{t-1}\sin\theta_t \\
x_2 &= \cos\theta_1 \sin\theta_2 \cdots \sin\theta_{t-1}\sin\theta_t \\
x_3 &= \cos\theta_2 \cdots \sin\theta_{t-1}\sin\theta_t \\
&\cdots \\
x_t &= \cos\theta_{t-1}\sin\theta_t \\
x_{t+1} &= \cos\theta_t\cos\theta_{t+1} \\
x_{t+2} &= \cos\theta_t \sin\theta_{t+1}\cos\theta_{t+2} \\
&\cdots \\
x_{n-1} &= \cos\theta_t \sin\theta_{t+1}\sin\theta_{t+2}\cdots\cos\theta_{n-1} \\
x_n &= \cos\theta_t \sin\theta_{t+1}\sin\theta_{t+2}\cdots\sin\theta_{n-1}
\end{aligned} \tag{1.25}$$

with the same representation for the y's in terms of parameters $\phi_1, \phi_2, \cdots \phi_{n-1}$. It is to be observed that in (1.24) and (1.25) $\Sigma x^2 = 1, \Sigma y^2 = 1$. This we may assume since q and z are invariant under such transformations.

In this new coordinate system, our samples $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ are taken as random points on a unit hypersphere about the origin in n dimensions. There is no loss of generality in this since x and y are assumed to be uncorrelated in the population and hence possess spherical symmetry of the density distribution in a space of n dimensions.

The element of probability for the x points on this hypersphere is proportional to the $(n-1)$-dimensional area on this sphere. Now the $n-1$ dimensional area is given by

$$\sqrt{g}\,d\theta_1\,d\theta_2\cdots d\theta_{n-1}$$

where g is a determinant of order $n-1$ in which the element in the i^{th} row and j^{th} column is

$$\sum_{\alpha=1}^{n} \frac{\partial x_\alpha \partial x_\alpha}{\partial \theta_i \partial \theta_j}.$$

When $i \neq j$, all these quantities vanish as can be seen by inspection from (1.25). When $i = j$, we have

$$\begin{aligned}
\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_1}\right)^2 &= \sin^2\theta_2 \sin^2\theta_3 \cdots \sin^2\theta_t \\
\sum_1^n \left(\frac{\partial x_2}{\partial \theta_2}\right)^2 &= \sin^2\theta_3 \cdots \sin^2\theta_t \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_t}\right)^2 &= 1 \\
\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{t+1}}\right)^2 &= \cos^2\theta_t \\
\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{t+2}}\right)^2 &= \cos^2\theta_t \sin^2\theta_{t+1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{n-1}}\right)^2 &= \cos^2\theta_t \sin^2\theta_{t+1} \cdots \sin^2\theta_{n-2}.
\end{aligned}$$

Therefore

$$g = \sin^2\theta_2 \sin^4\theta_3 \cdots \sin^{2(t-1)}\theta_t \cos^{2(n-t-1)}\theta_t \cdots \sin^2\theta_{n-2}$$

and hence the element of generalized area is given by

$$\sin\theta_2 \sin^2\theta_3 \cdots \sin^{t-1}\theta_t \cos^{n-t-1}\theta_t \sin^{n-t-2}\theta_{t+1} \cdots \sin\theta_{n-2}\, d\theta_1\, d\theta_2 \cdots d\theta_{n-1}. \tag{1.26}$$

Similarly we can show that the element of generalized area for the y point is

$$\sin\phi_2 \sin^2\phi_3 \cdots \sin^{t-1}\phi_t \cos^{n-t-1}\phi_t \sin^{n-t-2}\phi_{t+1} \cdots \sin\phi_{n-2}\, d\phi_1\, d\phi_2 \cdots d\phi_{n-1}. \tag{1.27}$$

The joint distribution of $\theta_1, \theta_2, \cdots, \theta_{n-1}$ and $\phi_1, \phi_2, \cdots, \phi_{n-1}$ (since the θ's are independent of the ϕ's) is proportional to the product of (1.26) and (1.27).

We now introduce four new sets of variables, u, v, u', v', defined by the following equations

$$x_i = u_i \sin\theta_i, \qquad y_i = v_i \sin\phi_i \qquad (i = 1, 2, \cdots, t) \tag{1.28}$$

$$x_j = u'_j \cos\theta_t \qquad y_j = v'_j \cos\phi_t \qquad (j = t+1, \cdots, n). \tag{1.29}$$

The u_i and v_i can be regarded as two points on a sphere in a space of t dimensions and u'_j and v'_j as two points on a sphere in a space of $n-t$ dimensions.

Let λ be the angle between the two points u and v and μ the angle between the two points u' and v'. Then

$$\cos\lambda = \sum_{i=1}^{t} u_i v_i ; \qquad \cos\mu = \sum_{j=t+1}^{n} u_j' v_j' .$$

The probability element for λ is proportional to $\sin^{t-2}\lambda\, d\lambda$, and that for μ is proportional to $\sin^{n-t-2}\mu\, d\mu$.

From the definition of u_i and v_i, we see that they depend only on $\theta_1, \theta_2, \cdots, \theta_{t-1}$; $\phi_1, \phi_2, \cdots, \phi_{t-1}$ respectively, and u_j' and v_j' depend only on $\theta_{t+1}, \theta_{t+2}, \cdots, \theta_{n-1}$; $\phi_{t+1}, \phi_{t+2}, \cdots, \phi_{n-1}$ respectively. It follows that the quantities λ, μ, θ_t, and ϕ_t are independently distributed.

The joint distribution of the θ's and ϕ's we integrate between constant limits with respect to all the variates except θ_t and ϕ_t. This gives for the joint distribution of θ_t and ϕ_t

$$A_n \sin^{t-1}\theta_t \sin^{t-1}\phi_t \cos^{n-t-1}\theta_t \cos^{n-t-1}\phi_t\, d\theta_t\, d\phi_t$$

where A_n is a constant depending only on n.

Multiplying this by the distributions of λ and μ and dropping the subscript t from θ and ϕ we get for the joint distribution of λ μ, θ, and ϕ

$$k_n \sin^{t-1}\theta \sin^{t-1}\phi \cos^{n-t-1}\theta \cos^{n-t-1}\phi \sin^{t-2}\lambda \cos^{n-t-2}\mu\, d\theta\, d\phi\, d\lambda\, d\mu \tag{1.30}$$

where k_n is a constant depending on n. The limits of integration for θ and ϕ are 0 and $\pi/2$; for λ and μ they are 0 and π.

Expressing q and z in terms of the new quantities as defined in (1.25), (1.28) and (1.29) we get

$$q^2 = \frac{\left(\sum_1^t x^2\right)\left(\sum_1^t y^2\right) - \left(\sum_1^t xy\right)^2}{1 - r^2} = \frac{\sin^2\theta \sin^2\phi \sin^2\lambda}{1 - r^2} \tag{1.31}$$

$$z = \frac{\left(\sum_{t+1}^n x^2\right)\left(\sum_{t+1}^n y^2\right) - \left(\sum_{t+1}^n xy\right)^2}{1 - r^2} = \frac{\cos^2\theta \cos^2\phi \sin^2\mu}{1 - r^2} \tag{1.32}$$

where

$$r = \Sigma xy = \sin\theta \sin\phi \cos\lambda + \cos\theta \cos\phi \cos\mu \tag{1.33}$$

is the sample correlation between x and y.

We now consider a transformation of the variables θ, ϕ, and μ in (1.30) to the new variables q, z, and r. Without troubling to compute the Jacobian J of the transformation, we know that it is independent of n since the relations (1.31), (1.32) and (1.33) do not involve n. Substituting from (1 31) and (1.32) into (1.30) we get for the joint distribution of q, z, r, and λ

$$k_n \psi q^{t-1} z^{\frac{1}{2}(n-t-1)} (1 - r^2)^{\frac{1}{2}(n-2)}\, dq\, dz\, dr\, d\lambda$$

where ψ is independent of n. Integrating with respect to λ between limits which are independent of n, we get for the joint distribution of q, z, and r

$$k_n \Psi q^{t-1} z^{\frac{1}{2}(n-t-1)} (1 - r^2)^{\frac{1}{2}(n-2)} \, dq \, dz \, dr. \tag{1.34}$$

But, for $n = t + 2$, this joint distribution reduces to (1.23) Therefore

$$k_{t+2} \Psi = \frac{1}{2} t(t-1) \frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}} z^{-1} q^{-1} (1 - r^2)^{-\frac{1}{2}}$$

so that (1.34) can be written as

$$k'_n q^{t-2} z^{\frac{1}{2}(n-t-3)} (1 - r^2)^{\frac{1}{2}(n-3)} \, dq \, dz \, dr.$$

However, since the distribution of r is known to be

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi}} (1 - r^2)^{\frac{1}{2}(n-3)} \, dr$$

we finally get for the joint distribution of q and z

$$h_n q^{t-2} z^{\frac{1}{2}(n-t-3)} \, dq \, dz$$

where h_n depends on n. The integral over the entire region defined by the inequalities

$$0 \leq q \leq 1, \qquad 0 \leq z \leq 1, \qquad z \leq (1 - q)^2$$

must equal unity; the constant h_n is therefore readily found to be $\frac{(n-2)!}{2(t-2)!\,(n-t-2)!}$. Thus the joint distribution in the final form is

$$\frac{(n-2)!}{2(t-2)!\,(n-t-2)!} q^{t-2} z^{\frac{1}{2}(n-t-3)} \, dq \, dz \,. \tag{1.35}$$

Now by (1.3) and (1.4), $q = r_1 r_2$, $z = (1 - r_1^2)(1 - r_2^2)$, and hence the Jacobian

$$\frac{\partial(q, z)}{\partial(r_1, r_2)} = 2(r_1^2 - r_2^2) \,. \tag{1.36}$$

Making the transformation in (1.35) we get the joint distribution of the canonical correlations r_1 and r_2 (for the case $s = 2$ and a general value of t) in the form

$$\frac{(n-2)!}{(t-2)!\,(n-t-2)!} (r_1^2 - r_2^2)(r_1 r_2)^{t-2} [(1 - r_1^2)(1 - r_2^2)]^{\frac{1}{2}(n-t-3)} \, dr_1 \, dr_2 \,. \tag{1.37}$$

II. JOINT LIMITING DISTRIBUTIONS OF CANONICAL CORRELATIONS AND LATENT ROOTS

In formula (1.37) we set

$$k_1 = nr_1^2, \qquad k_2 = nr_2^2$$

and get for the joint distribution of k_1 and k_2

$$\frac{(n-2)!}{4(t-2)!\,(n-t-2)!\,n^t}(k_1-k_2)(k_1k_2)^{\frac{1}{2}(t-3)} \left[\left(1-\frac{k_1}{n}\right)\left(1-\frac{k_2}{n}\right)\right]^{\frac{1}{2}(n-t-3)} dk_1\,dk_2. \tag{2.1}$$

When $n \to \infty$, the quantity $\dfrac{(n-2)!}{n^t(n-t-2)!}$ approaches 1 and $\left(1-\dfrac{k}{n}\right)^{\frac{1}{2}(n-t-3)}$ approaches $e^{-\frac{1}{2}k^2}$. Hence the limiting distribution of the two canonical correlations is given by

$$\frac{1}{4(t-2)!}(k_1-k_2)(k_1k_2)^{\frac{1}{2}(t-3)}\,e^{-\frac{1}{2}(k_1+k_2)}\,dk_1\,dk_2. \tag{2.2}$$

We shall call (2.2) the "generalized chi-square" distribution and show that the roots of the characteristic polynomial

$$\varphi(k) = \begin{vmatrix} a_{11}-k & a_{12} \\ a_{21} & a_{22}-k \end{vmatrix} \tag{2.3}$$

are distributed in precisely this form. Here $a_{ij} = \Sigma x_i x_j$ where x_1 and x_2 are normally and independently[6] distributed with unit variance in the population and zero mean in the sample.

Let k_1 and k_2 be the roots of (2.3). That is, k_1 and k_2 are the two roots of the quadratic equation

$$k^2 - p_1k + p_2 = 0 \tag{2.4}$$

where

$$p_1 = k_1 + k_2 = a_{11} + a_{22} \tag{2.5}$$

$$p_2 = k_1k_2 = a_{11}a_{22} - a_{12}^2. \tag{2.6}$$

In the absence of correlation in the population, the joint distribution of a_{11}, a_{22} and a_{12} is known to be

$$h_n \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}(a_{11}+a_{22})}\,da_{11}\,da_{22}\,da_{12} \tag{2.7}$$

where h_n is a constant depending only on n.

[6] The part of the assumption relating to independence may be removed without loss of generality. See last paragraph below.

We consider a transformation to the variables p_1, p_2 and a_{12}. From (2.5) and (2.6) we calculate the Jacobian J of the transformation,

$$J = \frac{1}{a_{11} - a_{22}} \tag{2.8}$$

and since

$$2a_{ii} = p_1 \pm (p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}} \tag{2.9}$$

$$J = \frac{1}{(p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}}}. \tag{2.10}$$

Substituting from (2.5) and (2.6) into (2.7) and multiplying by J, we get for the joint distribution of k_1, k_2 and a_{12}

$$h_n p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1} \frac{dp_1\, dp_2\, da_{12}}{(p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}}}. \tag{2.11}$$

We make the transformation $u = a_{12}^2$ and get for the joint distribution of k_1, k_2 and u

$$\frac{h_n}{2} p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1} \frac{dp_1\, dp_2\, du}{(bu - 4u^2)^{\frac{1}{2}}} \tag{2.12}$$

where $b = p_1^2 - 4p_2$.

Since both a_{11} and a_{22} are real, equation (2.9) shows that $b - 4u \geq 0$. Hence the limits of integration for u are 0 and $\frac{b}{4}$. Integrating out u in (2.12) between the above limits we obtain the joint distribution of p_1 and p_2.

Now the integral

$$\int_0^{b/4} \frac{du}{(bu - 4u^2)^{\frac{1}{2}}} = -\frac{1}{2} \sin^{-1}\left(\frac{-8u + b}{b}\right)\Bigg]_0^{b/4} = c \tag{2.13}$$

where c is some constant. Hence the joint distribution of p_1 and p_2 is given by

$$H_n p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1}\, dp_1\, dp_2. \tag{2.14}$$

By integrating (2.14) over the region $0 \leq p_2 \leq \left(\frac{p_1}{2}\right)^2$ and $0 \leq p_1 \leq \infty$ we get $H_n = \frac{1}{4}(n-2)!$.

We next transform p_1 and p_2 in terms of k_1 and k_2 from (2.5) and get for the joint distribution of k_1 and k_2

$$\frac{1}{4(n-2)!} (k_1 - k_2)(k_1 k_2)^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}(k_1+k_2)}\, dk_1\, dk_2. \tag{2.15}$$

This distribution is identical with that of (2.2) with $n = t$.

The above is an example of a more general

THEOREM: *Let $r_1, r_2, \cdots, r_s$ be a set of simple finite canonical roots of the two independent sets of variates $x_1, \cdots, x_s$, and $x_{s+1}, \cdots, x_{s+t}$. Let $k_i = nr_i^2$ $(i = 1, 2, \cdots, s)$. Then the joint limiting distribution of the k's approaches the exact joint sampling distribution of the latent roots of a matrix of sample product sums with t degrees of freedom of s normally distributed variates having unit variance in the population.*

PROOF: The proof follows from equation (1.11). For let us multiply and divide a_{ij} in (1.11) by n and set $n\lambda^2 = k$. The determinantal polynomial becomes

$$\varphi(k) = | b_{ij} - ks_{ij} |. \tag{2.16}$$

Without any loss of generality, we so transform the first set of variates that they become of zero correlation and unit variance in the population. Then it follows that

$$E(s_{ij}) = E\left(\sum \frac{x_i x_j}{n}\right) = \delta_{ij}$$

where δ_{ij} equals zero for $i \neq j$ and 1 for $i = j$.

Now let $P(x \geq a)$ stand for the probability that the variate x be greater than or equal to some constant a. Then, by the Strong Law of Large Numbers we can state that, given an $\epsilon > 0$ and a $\delta > 0$ there exists a positive integer n_0 such that for $n > n_0$

$$P\{| s_{ij} - \delta_{ij}| \geq \delta\} \leq \epsilon.$$

If then we let n increase indefinitely, the quantity $b_{ij} = \sum_1^t x_i x_j$ remains fixed while s_{ij} approaches, in the probability sense, δ_{ij}. Since the roots of a polynomial are continuous functions of the coefficients, we can, by an extension of the Law of Large Numbers, show that in the limit the roots of (2.16) will be distributed like the roots of the polynomial

$$\varphi(k) = | b_{ij} - k\delta_{ij} |.$$

This proves the theorem.

COROLLARY 1. *The limiting distribution of q^2 in case of complete independence between the two sets of variates approaches the exact distribution of a generalized sample variance (i.e. a determinant of sample variances and covariances) with t degrees of freedom. The proof follows from the fact that q^2 is a product of the roots of* (1.11) *and therefore by the above theorem, is distributed in the limit like* $| b_{ij} |$.

COROLLARY 2. *The distribution of the sum of the squares of the canonical correlations approach in the limit a χ^2 distribution with st degrees of freedom. This is obvious since in the limit the sum of the squares of the roots, by the above theorem, has the distribution of $b_{11} + b_{22} + \cdots + b_{ss}$ and each b_{jj} is distributed like χ^2 with t degrees of freedom.*

While the canonical roots of (1.2) are invariant under any non-singular linear transformations, the latent roots of a determinant of sample covariances are

invariant only under an orthogonal transformation. But there exists an orthogonal transformation which reduces a set of variates having a multivariate normal distribution to a set which are normally and independently distributed with variances equal to the latent roots of the population generalized variances of the original variates. Hence, in dealing with the distribution of latent roots, we may assume independence in the population without any loss of generality but the assumption of equal variance leads only to a special case. Moreover, the above consideration also explains the form of the asymptotic error of the sample latent root given in Part III of this paper.

III. ASYMPTOTIC STANDARD ERRORS OF LATENT ROOTS AND COEFFICIENTS OF PRINCIPAL COMPONENTS

1. Many statisticians have had occasion to use in their statistical analyses characteristic roots (or as they are sometimes called "latent" roots) of determinants of correlations or covariances. Especially has this become true since the publication of Hotelling's paper on principal components.[7] It is therefore of great importance to find, if not their sampling distributions, at least their limiting distributions and their asymptotic standard errors. This we shall do in this paper for the case of non-vanishing simple roots and by the same method[8] get the asymptotic variances and covariances of the coefficients of principal components. We have already derived in Part II the sampling distribution of the two latent roots of a determinant of covariances obtained from two normally distributed variates having equal variance in the population. This distribution is of no great importance in itself except that it gives us some idea as to the form of the distribution in the general case

In what follows, we shall use the convention that a repeated subscript in the same term stands for summation. If repeated subscripts appearing in a term are not to be summed, we shall place them in brackets following the expression in which they appear. Thus in the equation (3.1) below, we sum with respect to j but not with respect to q even though on the right hand side q appears twice.

Let $x_1, x_2, \cdots, x_n$ be a set of variates which have a multi-variate normal distribution. We assume that these variates have been resolved into components by Hotelling's method.[9] Let $\gamma_1, \gamma_2, \cdots, \gamma_n$ be the principal components. Then $x_i = a_{ij}\gamma_j$. The a_{ij}'s satisfy the following equations:

$$a_{jq}\sigma_{ij} = \lambda_q a_{iq}, \ [q] \tag{3.1}$$

$$a_{ip}a_{iq} = \lambda_q \delta_{pq} \tag{3.2}$$

---

[7] "Analysis of a Complex of Statistical Variables into Principal Components," *The Journal of Educational Psychology*, Sept. and Oct. 1933. See also M. A Girshick, "Principal Components," *Journal of the American Statistical Association*, Vol. 31, Sept. 1936.

[8] The method here employed is parallel to the one used by Hotelling in his paper of 1936 in deriving asymptotic standard errors for canonical correlations.

[9] Loc cit.

where the symbol δ_{pq} has the value zero for $p \neq q$ and 1 for $p = q$, λ_q is a root of the characteristic equation

$$| \sigma_{ij} - \lambda \delta_{ij} | = 0 \tag{3.3}$$

and σ_{ij} is the population covariance of x_i and x_j.

If we multiply (3.1) by a_{ip}, sum with respect to i and use (3.2), we get

$$a_{ip} a_{jq} \sigma_{ij} = \lambda_p^2 \delta_{pq} . \tag{3.4}$$

When a root of (3.3) is simple and not equal to zero, the corresponding a_{ij}'s and the root itself are definite analytic functions of the σ_{ij}'s over a region without singularities. A set of sampling errors $d\sigma_{ij}$ in the covariances will then determine a corresponding set of sampling errors in the a_{ij}'s and in the root.

We assume then, that the roots $\lambda_1, \lambda_2, \cdots, \lambda_t$ of (3.3) we are considering are simple and non-vanishing. In terms of the derivatives of the analytic functions we define

$$da_{rk} = \frac{\partial a_{rk}}{\partial \sigma_{pq}} d\sigma_{pq}, \qquad d\lambda_r = \frac{\partial \lambda_r}{\partial \sigma_{pq}} d\sigma_{pq} \tag{3.5}$$

where $d\sigma_{pq} = s_{pq} - \sigma_{pq}$, s_{pq} being the corresponding sample covariance.

Differentiating equation (3.1) and employing the above formulae we get

$$\sigma_{ij} da_{jq} + a_{jq} d\sigma_{ij} = \lambda_q da_{iq} + a_{iq} d\lambda_q . \qquad [q] \tag{3.6}$$

We now multiply this equation by a_{ip}, sum with respect to i, and use equations (3.1) and (3.2). This yields:

$$\lambda_p a_{jp} da_{jq} + a_{ip} a_{jq} d\sigma_{ij} = \lambda_q a_{ip} da_{iq} + \lambda_q \delta_{pq} d\lambda_q . \qquad [p, q] \tag{3.7}$$

When $p = q$, the term $\lambda_p a_{jp} da_{jp}$ cancels out and equation (3.7) reduces to

$$\lambda_p d\lambda_p = a_{ip} a_{jp} d\sigma_{ij} . \qquad [p] \tag{3.8}$$

We change the subscripts p, i, j, to q, k, m, in (3.8) and multiply together the two equations thus obtained. This gives:

$$\lambda_p \lambda_q d\lambda_p d\lambda_q = a_{ip} a_{jp} a_{kq} a_{mq} d\sigma_{ij} d\sigma_{km} . \qquad [p, q] \tag{3.9}$$

Hence

$$\lambda_p \lambda_q E(d\lambda_p d\lambda_q) = a_{ip} a_{jp} a_{kq} a_{mq} E(d\sigma_{ij} d\sigma_{km}) \qquad [p, q] \tag{3.10}$$

where the symbol E denotes the mathematical expectation or mean value of the expression following.

Now it can be easily shown by means of the characteristic function of a multivariate normal distribution that

$$E(d\sigma_{ij} d\sigma_{km}) = \frac{1}{n} (\sigma_{ik} \sigma_{jm} + \sigma_{im} \sigma_{jk}) \tag{3.11}$$

where n is one less than the number in the sample. Substituting this expression in equation (3.10) and using (3.4) we get the following rather simple result

$$\lambda_p\lambda_q E(d\lambda_p d\lambda_q) = \frac{2\lambda_p^4\delta_{pq}}{n}. \qquad [p, q] \tag{3.12}$$

Setting $p = q$ in this formula we get

$$E[(d\lambda_p)^2] = \frac{2\lambda_p^2}{n}. \tag{3.13}$$

But when $p \neq q$

$$E[d\lambda_p d\lambda_q] = 0. \tag{3.14}$$

Let $l_1, l_2, \cdots, l_t$, be the corresponding latent roots of a determinant of sample covariances. The sample latent root l_p may be expanded about λ_p in a Taylor series of the form

$$l_p = \lambda_p + \frac{\partial\lambda_p}{\partial\sigma_{rl}} d\sigma_{rl} + \frac{1}{2}\frac{\partial^2\lambda_p}{\partial\sigma_{rl}\partial\sigma_{uv}} d\sigma_{rl}d\sigma_{uv} + \cdots \tag{3.15}$$

or, by (3.5)

$$l_p - \lambda_p = d\lambda_p + \cdots. \tag{3.16}$$

Squaring both sides of (3.16), taking the expected value, and using (3.13) we find *that the sample variance of a latent root l_p, apart from terms of higher order in n^{-1}, is given by* $\frac{2\lambda_p^2}{n}$.

If in (3.11) we set $i = j = k = m$, we get the variance of a sample variance, and it is interesting to note that its form is identical with the first term of the asymptotic expansion of the variance of a sample latent root.

The sample covariance of any two distinct roots is by (3.14) zero for the first term of the asymptotic expansion. That is, the covariance is at least of order n^{-2}. All the above results also follow from the fact, shown by the author in a previous paper,[10] that the coefficients of the principal components and hence the latent roots are maximum likelihood statistics. This property of the latent roots permits us also to state the following

THEOREM: *Let $\lambda_1, \lambda_2, \cdots, \lambda_t$ be any set of simple non-vanishing roots of* (3.3). *For sufficiently large samples these will be approximated by certain of the latent roots $l_1, l_2, \cdots, l_t$ of the samples. If $l_i - \lambda_i$ is divided by the standard error*

$$\sigma_{l_i} = \lambda_i\sqrt{\frac{2}{n}}$$

the resulting variates have a distribution which, as n increases, approaches the normal distribution of t independent variates of zero mean and unit standard deviation.

[10] Loc. cit

COROLLARY: Let λ_1 be a maximum simple, non-vanishing root of (3 3) and let l_1 be the corresponding maximum sample root. Then, $l_1 - \lambda_1$ divided by its standard error has a distribution approaching normality in the limit.

2. The Variance of Log l. The formula for the standard error of the latent root given above contains a population parameter λ the numerical value of which we usually do not know. It is therefore important to find a transformation of the latent root to a new variate which will have or its leading term of the asymptotic standard error a quantity independent of the population parameter.

Let $k = f(l)$ be such a transformation. Then $K = f(\lambda)$ is the corresponding transformation for the population root.

We now expand k in a Taylor series about $l = \lambda$.

$$dk = f'(\lambda)dl + \tfrac{1}{2}f''(\lambda)(dl)^2 + \cdots \tag{3.17}$$

and get an approximation

$$dk = f'(\lambda)dl. \tag{3.18}$$

Squaring both sides and taking the expectation, we get

$$E(dk)^2 = [f'(\lambda)]^2 E[(dl)^2] = [f'(\lambda)]^2 \frac{2\lambda^2}{n}. \tag{3.19}$$

Now set $E(dk)^2 = 2/n$. Then, from (3.19)

$$f'(\lambda) = 1/\lambda$$

or

$$f(\lambda) = \log \lambda \tag{3 20}$$

Hence, if we set $k = \log l$, then

$$\sigma_k^2 = 2/n \tag{3.21}$$

is an approximation to the variance of k and is independent of any population parameter.

3. The Asymptotic Variances and Covariances of Roots of Determinants of Correlations. While the formulas for the asymptotic standard errors of the latent roots of a determinant of covariances are rather simple, this is not the case with the roots of a determinant of correlations. In deriving the asymptotic standard errors of simple non-vanishing roots of a determinant of correlations, we again assume that the variates $x_1, x_2, \cdots, x_s$, which in this case are of unit variance in the population, have been resolved into principal components. The equations of the previous section, up to and including (3.10), remain the same except that we substitute ρ_{ij} for every σ_{ij}, where ρ_{ij} is the population correlation of x_i with x_j. Thus equation (3.10) becomes

$$\lambda_p\lambda_q E(d\lambda_p d\lambda_q) = a_{ip}a_{jp}a_{kq}a_{mq}E(d\rho_{ij}d\rho_{km}), \qquad [p, q] \tag{3.22}$$

where $d\rho_{ij} = r_{ij} - \rho_{ij}$, r_{ij} being the sample correlation between x_i and x_j. The expected value of $d\rho_{ij}d\rho_{km}$ is not, as in the case of the σ's given in the simple form of (3.11) but rather it is given asymptotically, the leading term in n^{-1} being the following lengthy expression:

$$\begin{aligned} nE(d\rho_{ij}d\rho_{km}) &= \rho_{ik}\rho_{mj} + \rho_{kj}\rho_{mi} - \rho_{ij}\rho_{ki}\rho_{mi} - \rho_{ij}\rho_{kj}\rho_{mj} \\ &\quad - \rho_{km}\rho_{ki}\rho_{kj} + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{ki}^2 + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{kj}^2 \\ &\quad - \rho_{km}\rho_{mi}\rho_{mj} + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{mi}^2 + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{mj}^2 . \end{aligned} \tag{3.23}$$

$[i, j, k, m]$

Substituting this in (3.22) and simplifying by means of equations (3.1) and (3.4) we finally get

$$\begin{aligned} n\lambda_p\lambda_q E(d\lambda_p d\lambda_q) &= 2(\lambda_p^4\delta_{pq} + \lambda_p\lambda_q a_{ip}^2 a_{jq}^2\rho_{ij}^2) \\ &\quad - 2(\lambda_p\lambda_q^2 a_{ip}^2 a_{iq}^2 + \lambda_p^2\lambda_q a_{jp}^2 a_{jq}^2). \end{aligned} \tag{3.24}$$

$[p, q]$

When $p = q$, (3.24) becomes

$$E[(d\lambda_p)^2] = \frac{2}{n}\left[\lambda_p^2 + a_{ip}^2 a_{jp}^2\rho_{ij}^2 - 2\lambda_p\sum_{i=1}^{s} a_{ip}^4\right]. \tag{3.25}$$

$[p]$

When $p \neq q$,

$$E(d\lambda_p d\lambda_q) = \frac{2}{n}[a_{ip}^2 a_{jq}^2\rho_{ij}^2 - (\lambda_p + \lambda_q)a_{ip}^2 a_{iq}^2]. \tag{3.26}$$

$[p, q]$

Hence (3.25) is the leading term of the asymptotic expansion of the variance of λ_p, and (3.26) is the leading term of the asymptotic expansion of the covariance of λ_p and λ_q, where λ_p and λ_q are simple, non-vanishing roots of a determinant of correlations.

4. **Asymptotic Variances and Covariances of Coefficients of Principal Components Derived from a Determinant of Covariances.** Let $x_i = a_{ij}\gamma_j$ be the equation of transformation of the variates $x_1, x_2, \cdots, x_s$ into principal components. In what follows we assume that the latent roots of the determinantal equation (3.3) are simple and none equal to zero. The last restriction makes the determinant of covariances non-vanishing. The determinant of the a_{ij}'s will therefore be also different from zero. With these assumptions in mind, we now proceed to derive the asymptotic variances and covariances of the a_{ij}'s.

We set $p = q$ in (3.2) and differentiate the result. This yields:

$$d\lambda_p = 2a_{lp}da_{lp}, \tag{3.27}$$

$[p]$

where the summation index i was replaced by l. Substituting for $d\lambda_p$ from (3.8) we get:

$$a_{ip}a_{jp}d\sigma_{ij} = 2\lambda_p a_{lp}da_{lp}. \tag{3.28}$$

$[p]$

Now, when $p \neq q$, equation (3.7) reduces to

$$\lambda_p a_{jp}da_{jq} + a_{ip}a_{jq}d\sigma_{ij} = \lambda_q a_{ip}da_{iq},$$

$[p, q]$

or

$$a_{ip}a_{jq}d\sigma_{ij} = (\lambda_q - \lambda_p)a_{lp}da_{lq}\,. \qquad [p, q] \tag{3.29}$$

We combine equations (3.28) and (3.29) into one equation

$$a_{ip}a_{jq}d\sigma_{ij} = (\lambda_q + \epsilon_{pq}\lambda_p)a_{lp}da_{lq}\,, \qquad [p, q] \tag{3.30}$$

where ϵ_{pq} has the value 1 when $p = q$ and -1 when $p \neq q$. The reciprocal of $\lambda_q + \epsilon_{pq}\lambda_p$, (which is different from zero), we denote by b_{qp}. Then equation (3.30) can be written as

$$a_{ip}b_{qp}a_{jq}d\sigma_{ij} = a_{lp}da_{lq}\,. \qquad [p, q] \tag{3.31}$$

Since the determinant $|\, a_{ij} \,|$ of the a_{ij}'s is different from zero, we can solve this set of homogeneous linear equations for da_{lq}'s, $(l = 1, 2, \cdots, s)$. To do this we multiply equation (3.31) by A^{tp}, where A^{tp} is the element of the t^{th} row and p^{th} column of the inverse of the determinant $|\, a_{ij} \,|$, and sum with respect to p. Since $A^{tp}a_{lp} = \delta_{tl}$ we get,

$$A^{tp}a_{ip}b_{qp}a_{jq}d\sigma_{ij} = \delta_{tl}da_{lp} = da_{tq}\,. \qquad [q] \tag{3.32}$$

We now change the subscripts i, j, t, p, q, in (3.32) to k, m, r, u, v, respectively, multiply the two equations thus obtained, and take the expected value:

$$E(da_{tq}da_{rv}) = A^{tp}A^{ru}a_{ip}a_{ku}b_{qp}b_{vu}a_{jq}a_{mv}E(d\sigma_{ij}d\sigma_{km}). \qquad [q, v] \tag{3.33}$$

Substituting for $Ed\sigma_{ij}d\sigma_{km}$ its values from (3.11) and simplifying by means of (3.4) we get:

$$E(da_{tq}da_{rv}) = \frac{\lambda_v^2\lambda_q^2}{n}A^{tv}A^{rq}b_{vq}b_{qv} + \frac{\lambda_q^2\delta_{qv}}{n}\sum_{u=1}^{s}A^{tu}A^{ru}b_{qu}b_{vu}\lambda_u^2 \tag{3.34}$$

where we sum *only with respect to* u We may simplify this formula to some extent by employing the relation: $A^{tq} = a_{tq}/\lambda_q$. (This relation is obtained from (3.2) by multiplying each side of that equation by A^{tp} and summing with respect to p). When this is done and the values for the b's are substituted, the final result becomes:

$$\begin{aligned} E(da_{tq}da_{rv}) &= \frac{\lambda_v\lambda_q a_{tv}a_{rq}}{n(\lambda_q + \epsilon_{qv}\lambda_v)(\lambda_v + \epsilon_{qv}\lambda_q)} \\ &\quad + \frac{\lambda_q^2\delta_{qv}}{n}\sum_{u=1}^{s}\frac{a_{tu}a_{ru}}{(\lambda_q + \epsilon_{qu}\lambda_u)(\lambda_v + \epsilon_{vu}\lambda_u)}. \end{aligned} \tag{3.35}$$

From this we derive the following specific formulas:

$$\begin{aligned} E(da_{tq}da_{rq}) &= \frac{a_{tq}a_{rq}}{4n} \\ &\quad + \frac{\lambda_q^2}{n}\left[\frac{a_{t1}a_{r1}}{(\lambda_q - \lambda_1)^2} + \cdots + \frac{a_{tq}a_{rq}}{4\lambda_q^2} + \cdots + \frac{a_{ts}a_{rs}}{(\lambda_q - \lambda_s)^2}\right] \end{aligned} \tag{3.36}$$

$$(3.37)\qquad E[(da_{tq})^2] = \frac{a_{tq}^2}{4n} + \frac{\lambda_q^2}{n}\left[\frac{a_{t1}^2}{(\lambda_q - \lambda_1)^2} + \cdots + \frac{a_{tq}^2}{4\lambda_q^2} + \cdots + \frac{a_{ts}^2}{(\lambda_q - \lambda_s)^2}\right]$$

$$(3.38)\qquad E(da_{tq}\,da_{rv}) = -\frac{\lambda_q\lambda_v a_{tv}a_{rq}}{n(\lambda_q - \lambda_v)^2}. \qquad (q \neq v)$$

Formulas (3.36), (3.37) and (3.38) give us the leading terms of the asymptotic expansions of the variances and covariances for the principal components. It should be remarked that the coefficients of "mutual regression" equations can be easily shown to be proportional to those of the principal components. Hence their asymptotic standard errors and covariances may be derived in a similar manner and will be of the same form.

5. **Variances and Covariances of Latent Roots when the Population Roots are Equal.** Let $k_1, k_2, \cdots, k_p$ be the latent roots of a generalized sample variance of p normally distributed variates.

Ordinarily the subscripts of the roots designate their ranks, so that $k_1 \geq k_2 \geq \cdots \geq k_p$. We may, however, assign to a root a subscript from 1 to p without any regard to its size.[11] If this is done randomly for every sample of n observations the mathematical expectation of $k_i^r k_j^s k_k^t \cdots$ will be the same for every permutation of the subscripts $i, j, k, \cdots$. This fact permits us to calculate the variances and covariances of the above roots.

We may assume, without any loss of generality, that the p variates are independently distributed,[12] and furthermore we assume the population roots to be all equal to unity. Then equation (3.11) becomes

$$(3.39)\qquad E(s_{ij}s_{km}) = \delta_{ij}\delta_{km} + \frac{1}{n}(\delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}).$$

Where s_{pq} is the sample variance of x_p and x_q and δ_{pq} is the Kronecker delta.

Now it can be easily shown that

$$(3.40)\qquad \sum_1^p s_{ii} = \sum_1^p k, \quad \sum_{i<j}(s_{ii}s_{jj} - s_{ij}^2) = \sum_{i<j} k_i k_j, \quad \sum_1^p s_{ii}^2 + 2\sum_{i<j} s_{ij}^2 = \sum_1^p k^2.$$

Hence $E(k) = 1$, and

$$E(\sum k^2) = E(\sum s_{ii}^2 + 2\sum_{i<j} s_{ij}^2)$$

or

$$(3.41)\qquad pEk^2 = pEs_{ii}^2 + p(p-1)Es_{ij}^2. \qquad (i \neq j)$$

Substituting from (3.39) in (3.41) we get

$$E(k^2) = 1 + \frac{p+1}{n}.$$

[11] This approach was suggested to the author by Professor Hotelling.

[12] See Part II, last Paragraph

The variance of k is therefore given exactly by

$$\sigma_k^2 = E(k^2) - 1 = \frac{p_{\cdot\cdot} + 1}{n}. \tag{3.42}$$

In a similar manner we find the covariances of k_i and k_j to be

$$\sigma_{k_i k_j} = -\frac{1}{n} \tag{3.43}$$

IV. DISTRIBUTION AND MOMENTS OF QUANTITIES RELATED TO q AND z

From the known distribution of q and z and their expressions in terms of the ratio of determinants given by (1.1) and (1.12), we can derive moments and distributions of several related functions of sample variances and correlations of two independent sets of variates.

$$\text{Let } p = \frac{q^2}{z} = \frac{|b_{ij}|}{|c_{ij}|} \text{ by (1.12).} \tag{4.1}$$

Since the two determinants in (4.1) are independently distributed, the sampling distribution of p, given in the above form, can be obtained for a general value of s and t from Wilks'[13] distribution of the ratio of independent generalized variances.

Thus, for $s = 2$ and $t \geq 2$, the distribution of p is given by

$$\frac{\Gamma(n-2)}{2\Gamma(t-1)\Gamma(n-t-1)} p^{\frac{1}{2}(t-3)} \frac{dp}{(1+\sqrt{p})^{n-2}}. \tag{4.2}$$

When the number of variates in each set is the same, the numerator of q^2 in (1.1) becomes the square of the determinant of covariances *between* the two sets of variates. Thus

$$q^2 = \frac{|a_{i\alpha}|^2}{|a_{ij}||a_{\alpha\beta}|} \tag{4.3}$$

where $i, j,$ take on values from 1 to s, α, β take on values from $s+1$ to $2s$, and $a_{uv} = \sum_1^n x_u x_v$.

If the two sets are independent, the quantities q^2, $|a_{ij}|$, $|a_{\alpha\beta}|$, are independently distributed. Hence

$$E(|a_{i\alpha}|^m) = Eq^m(|a_{ij}|^{\frac{1}{2}m})\, E(|a_{\alpha\beta}|^{\frac{1}{2}m}). \tag{4.4}$$

Setting $\beta = 0$ in (1.16) and employing formula (1.15) we get for the moment of $|a_{i\alpha}|$

$$E(|a_{i\alpha}|^m) = \frac{2^{sm}}{|A_{ij}|^{\frac{1}{2}m}|A_{\alpha\beta}|^{\frac{1}{2}m}} \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{s+m-i+1}{2}\right)\Gamma\left(\frac{n+m-i+1}{2}\right)}{\Gamma\left(\frac{s-i+1}{2}\right)\Gamma\left(\frac{n-i+1}{2}\right)} \right] \tag{4.5}$$

[13] Loc. cit., pp. 478–479.

where A_{uv} denotes the cofactor corresponding to σ_{uv} divided by the determinant $|\sigma_{uv}|$, σ_{uv} being the population covariance of x_u and x_v.

We may replace the product sums in (4.3) by sample correlations and, with the assumption that all the variates come from independent populations, obtain the m^{th} moment of the determinant of correlations between the two sets as

$$(4.6)\quad E(|r_{i\alpha}|^m) = \frac{\Gamma^{2s}\left(\frac{n}{2}\right)}{\Gamma^{2s}\left(\frac{n+m}{2}\right)} \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{n+m-i+1}{2}\right)\Gamma\left(\frac{s+m-i+1}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)\Gamma\left(\frac{s-i+1}{2}\right)} \right].$$

This follows from the expression for the m^{th} moment of q and the formula

$$(4.7)\quad E(|r_{uv}|^k) = \frac{\Gamma^{t}\left(\frac{n}{2}\right)}{\Gamma^{t}\left(\frac{n+2k}{2}\right)} \prod_{i=1}^{t} \left[\frac{\Gamma\left(\frac{n+2k-i+1}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)} \right]$$

derived by Wilks.[14]

If we set $s = t = 2$, the numerator of q^2 in (4.3) becomes the square of a determinant of sample covariances (or correlations) known to psychologists as the tetrad. We shall here derive its distribution under the assumptions that the four variates are independently distributed.

We write

$$(4.8)\quad q = \frac{T}{u_1 u_2}$$

where

$$(4.9)\quad T = r_{13}r_{24} - r_{14}r_{23}, \qquad u_1 = (1 - r_{12}^2)^{\frac{1}{2}}, \qquad u_2 = (1 - r_{34}^2)^{\frac{1}{2}}$$

and q is taken as positive.

Now the distribution of q for $s = t = 2$ is given by

$$(4.10)\quad (n-2)(1-q)^{n-3}dq$$

and the distribution of u is known to be

$$(4.11)\quad \frac{2\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} u^{n-2}(1-u^2)^{-\frac{1}{2}}\,du.$$

Hence the distribution of u_1, u_2 and q is given by

$$(4.12)\quad \frac{4(n-2)}{\pi} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} (1-q)^{n-3}(u_1u_2)^{n-2}[(1-u_1^2)(1-u_2^2)]^{-\frac{1}{2}}\,du_1\,du_2\,dq$$

[14] Loc. cit., p. 492.

Performing the transformation (4.8) and integrating out u_1 and u_2 we get for the distribution of the tetrad

$$(4.13) \qquad \frac{4(n-2)\Gamma^2\left(\frac{n}{2}\right)}{\pi\Gamma^2\left(\frac{n-1}{2}\right)} \int_T^1 \int_{\frac{T}{u_1}}^1 \frac{(u_1u_2 - T)^{n-3}}{\sqrt{(1-u_1^2)(1-u_2^2)}}\, du_1, du_2.$$

All the moments of T can of course be obtained by setting $s = 2$ in (4.6).[15]

U. S. Department of Agriculture,
Washington, D. C.

[15] The limiting distribution of the tetrad was given by J. L Doob in an article entitled "The Limiting Distributions of Certain Statistics," *Annals of Mathematical Statistics*, Vol. 6, (1935). For a more general distribution of the tetrad and other statistics considered in this paper see W. G. Madow, "Contributions to the Theory of Multivariate Statistical Analysis," *Transactions of the American Mathematical Society*, Nov. 1938.

AN OPTIMUM PROPERTY OF CONFIDENCE REGIONS ASSOCIATED WITH THE LIKELIHOOD FUNCTION[1]

By S. S. Wilks and J. F. Daly

One of the authors [1] has recently established a connection between the method of maximum likelihood and shortest average confidence intervals for the case of one unknown parameter, and has reported a generalization [2] of this result for the case of several parameters. It is the object of this paper to consider the several-parameter problem in greater detail and at the same time to make the previously obtained result slightly stronger, particularly in the one-parameter case.

Let x denote a set of random variables, and θ a set of parameters $\theta_1, \cdots, \theta_h$. Suppose Π_0 is a population with the cumulative distribution function $F(x, \theta_0) \equiv F_0$ say. Then the logarithm of the likelihood associated with the population Π_0^n of random samples $0_n: x_1, x_2, \cdots, x_n$ drawn from Π_0 is

$$L^n(x, \theta_0) = \sum_{\alpha=1}^{n} \log dF(x_\alpha, \theta_0).$$

For a given sample 0_n we shall say that a set of functions $H_i^n(x, \theta)$ is of class K if there exists a domain R of parameter points θ: $(\theta_1, \cdots, \theta_h)$ in a θ-space such that for each θ_0 in R:

(i) $H_i^n(x, \theta_0) = H_{i0}^n$ is of the form $\sum_{\alpha=1}^{n} h_i(x_\alpha, \theta_0)$;

(ii) $h_i(x, \theta_0) = h_{i0}$ exists for all x except possibly for a set of zero probability;

(iii) $E_0[h_{i0}] = 0$, where E_0 means that the expected value is taken for the population Π_0;

(iv) $\| E_0[h_{i0}h_{j0}] \|$ exists and is non-singular;

(v) the moments $E_0[h_{i0}h_{j0}h_{k0}]$ are all finite.

(Here and throughout the remainder of the paper, the indices i, j, k, l have the range $1, \cdots, h$.) If, in addition,

(iii′) $E_0[h_{i0}]$ can be differentiated under the integral sign;

(iv′) the moments $E_0[h_{i0}h_{j0}]$ are differentiable with respect to the θ's;

the H_i will be said to be of class K'.

We shall need the following lemma, which is very closely related to Theorem 1′ and Theorem 2 in [1] and which can be proved by the method of characteristic functions.

---

[1] Incorporated in this paper is a note presented by one of us (c.f. [2]) at a meeting of the Institute of Mathematical Statistics, December 27, 1938.

LEMMA: *Let $H_i^n(x, \theta)$ be of class K for each n, and put*

$$B_{ij0}^n = \frac{1}{n} E_0[H_{i0}^n H_{j0}^n] = E_0[h_{i0} h_{j0}].$$

Let $\| b_{ij0}^n \|$ be the positive definite matrix satisfying the equation

$$\| b_{ij0}^n \|^2 = \| B_{ij0}^n \|$$

and write

$$\| b_{ij0}^n \|^{-1} = \| b_0^{nij} \|$$

Then for any point θ_0 in R the functions

$$\varphi_{i0}^n = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} b_0^{nij} H_{j0}^n \tag{1}$$

computed from Π_0^n have a joint distribution which converges in large samples to normality, with the density function

$$(2\pi)^{-\frac{k}{2}} e^{-\frac{1}{2}\sum_{i=1}^{k} \varphi_{i0}^2}$$

Now whenever we are justified in assuming a definite functional form for $F(x, \theta)$, and have a set of functions $\varphi_i(x, \theta)$ whose distribution under this last assumption is known and is independent of the θ's, as is the case in the limit for the functions (1), we can obtain, from a sample, information about the values of the θ's. For, given any region S in the space of the functions φ_i, we can determine the probability $P_0\{\varphi_{i0} \subset S\}$ that in samples from Π_0 the point $(\varphi_{10}, \cdots, \varphi_{k0})$ will fall in the region S, even though we do not know the population values θ_0. Suppose, then, that we pick a region S such that $P_0\{\varphi_{i0} \subset S\} > .95$, and agree that each time we encounter such a problem we shall substitute the observed x's into the φ's, and call the set of all points $(\theta_1, \cdots, \theta_k)$ for which $\varphi_i(x, \theta) \subset S$ the *confidence region T*. If this procedure is followed consistently, we can assert that the probability is more than .95 that the region T thus determined contains the true parameter point θ_0.

Evidently the size of the confidence region, i.e, the accuracy with which it serves to locate the true parameter point θ_0, depends upon our choice of the auxiliary functions φ_i. Consider now the case in which there is but one parameter θ, and let $\varphi(x, \theta)$ and $\varphi^*(x, \theta)$ be two functions with the same distribution $D(u)$, where $D(u)$ does not depend on θ. For the set S of the above discussion take the interval $\underline{u} < u < \bar{u}$. Then

$$P_0\{\varphi_0 \subset S\} = P_0\{\varphi_0^* \subset S\} = \alpha$$

where $\alpha = .95$, say. Given a set of observed x's, $\varphi(x, \theta)$ will map S into a confidence region T, while $\varphi^*(x, \theta)$ will map it into a confidence region T^*. Both T and T^* may be expected to contain the true value θ_0 in 95% of the cases; hence a reasonable way to compare the size of T with that of T^* is to compare the

quantities $\frac{\partial \varphi}{\partial \theta}(x, \theta_0)$ and $\frac{\partial \varphi^*}{\partial \theta}(x, \theta_0)$; for these derivatives give an indication of the amount of change one can make in θ without forcing φ or φ^* out of the interval S.

The result obtained in [1] in this connection may now be stated as follows: Let $H = \frac{\partial L}{\partial \theta}$ be of class K', and let $H^* = \sum_{\alpha=1}^{n} h(x_\alpha, \theta)$ be any other function of class K'. Then in large samples from Π_0 both

$$\varphi = \frac{H}{\left(nE\left[\left\{\frac{\partial}{\partial \theta} \log dF\right\}^2\right]\right)^{\frac{1}{2}}}$$

and

$$\varphi^* = \frac{H^*}{(nE[\{h(x, \theta)\}])^{\frac{1}{2}}}$$

are distributed almost normally with zero mean and unit variance But the confidence regions obtained from φ will, on the average, be smaller than those from φ^*, in the sense that, for large samples the inequality

$$\left\{E_0\left[\frac{\partial \varphi_0}{\partial \theta}\right]\right\}^2 > \left\{E_0\left[\frac{\partial \varphi_0^*}{\partial \theta}\right]\right\}^2 \tag{2}$$

will hold (unless $h(x, \theta) \equiv c \frac{\partial}{\partial \theta} \log dF$, in which case alone the inequality (2) becomes an equality).

Now let us return to the several-parameter case. One method of attack which suggests itself is to consider the jacobian determinant

$$\left|\frac{\partial \varphi_{i0}}{\partial \theta_j}\right|$$

for this bears the same relation to the area of the region dS which maps into the region

$$dT\colon\ \theta_0 - \tfrac{1}{2}d\theta < \theta < \theta_0 + \tfrac{1}{2}d\theta$$

as does the derivative $\frac{\partial \varphi_0}{\partial \theta}$ in the one parameter case. To this end, let us put $L_i^n(x, \theta) = \frac{\partial L^n}{\partial \theta_i}$, and for each n and for each θ_0 in R assume that

(a) L_{i0}^n is defined for all x except perhaps on a set of probability 0;
(b) $E_0[L_{i0}^n] = 0$;
(c) $E_0[L_{i0}^n]$ can be differentiated under the integral sign;
(d) $\| E_0[L_{i0}^n L_{j0}^n] \|$ exists and is non-singular;
(e) $E_0[L_{i0}^n L_{j0}^n]$ is differentiable in the θ's

Let $H_i^n(x, \theta)$ be any other set of functions satisfying the same conditions Set

$$E_0[L_{i0}^n L_{j0}^n] = nA_{ij0}^n \qquad E_0[H_{i0}^n H_{j0}^n] = nB_{ij0}^n$$

and define the matrices

$$\| a_{ij0}^n \|^2 = \| A_{ij0}^n \| \qquad \| a_0^{nij} \| = \| a_{ij0}^n \|^{-1}$$
$$\| b_{ij0}^n \|^2 = \| B_{ij0}^n \| \qquad \| b_0^{nij} \| = \| b_{ij0}^n \|^{-1}$$

Now consider the normalized functions

$$\tilde{L}_{i0}^n = \sum_{j=1}^{h} a_0^{nij} L_{j0}^n$$

$$\tilde{H}_{i0}^n = \sum_{j=1}^{h} b_0^{nij} H_{j0}^n$$

We then have

$$\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k} = \sum_{j=1}^{h} \frac{\partial a_0^{nij}}{\partial \theta_k} \cdot \frac{1}{n} \cdot L_{j0}^n + \sum_{j=1}^{h} a_0^{nij} \cdot \frac{1}{n}\frac{\partial L_{j0}^n}{\partial \theta_k} \tag{3}$$

and by virtue of assumptions (b) and (c) it follows that (c.f. [1], pp. 171–2)

$$E_0\left[\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k}\right] = -\frac{1}{n}\sum_{j=1}^{h} a_0^{nij} E_0[L_{j0}^n L_{k0}^n]$$

In similar fashion

$$E_0\left[\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial \theta_k}\right] = -\frac{1}{n}\sum_{j=1}^{h} b_0^{nij} E_0[H_{j0}^n L_{k0}^n]$$

Consequently

$$(-1)^h \left| E_0\left[\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k}\right] \right| = | A_{ij0}^n |^{\frac{1}{2}} \tag{4}$$

and

$$(-1)^h \left| E_0\left[\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial \theta_k}\right] \right| = | B_{ij0}^n |^{-\frac{1}{2}} \cdot \left| \frac{1}{n} E_0[H_{i0}^n L_{j0}^n] \right| \tag{5}$$

We can find a relation between these two determinants by going over to the matrix

$$M_n = \begin{Vmatrix} \| E_0[L_{i0}^n L_{j0}^n] \| & \| E_0[L_{i0}^n H_{j0}^n] \| \\ \| E_0[H_{i0}^n L_{j0}^n] \| & \| E_0[H_{i0}^n H_{j0}^n] \| \end{Vmatrix}$$

This matrix is positive definite unless there is a linear relation with constant coefficients, say $\sum (c_i L_i + d_i H_i) = 0$, which holds for all x's except a set of zero probability; and in this event it is positive semidefinite. From the theory of compound matrices [3] we can then conclude that the matrix whose elements are the h-th order minors of M_n arranged in lexicographic order on both row and column indices has the same property, so that

$$| E_0[L_{i0}^n L_{j0}^n] | \cdot | E_0[H_{i0}^n H_{j0}^n] | \geq | E_0[L_{i0}^n H_{j0}^n] |^2$$

The relations (4) and (5) then imply that

$$\left|\det E_0\left[\frac{1}{n}\frac{\partial \bar{L}^n_{i0}}{\partial \theta_k}\right]\right| \geq \left|\det E_0\left[\frac{1}{n}\frac{\partial \bar{H}^n_{i0}}{\partial \theta_k}\right]\right| \tag{6}$$

It may be observed that no use has been made of the assumption of linearity (i) in deriving (6). And since in the one parameter case the determinants have but one row and column, we see that in this case the result in [1] remains valid for functions of an even more general type than those of class K'. In order to give the inequality a statistical meaning it seems necessary, however, to require not only that H and L satisfy (a), $\cdots$ (e) but also that in large samples $\frac{1}{\sqrt{n}}\bar{H}^n_1$ and $\frac{1}{\sqrt{n}}\bar{L}^n_1$ tend to be distributed independently of θ, with the same (though not necessarily normal) distribution.

For the case of several parameters the transition from the above determinants of expected values to the jacobian determinants requires further argument and further assumptions. To begin with, suppose that the L^n_i and H^n_i are of class K', and that

(vi) the moments $E_0\left[\frac{\partial h_{i0}}{\partial \theta_j}\frac{\partial h_{k0}}{\partial \theta_l}\right]$ are all finite,

with a corresponding condition on the variances and covariances of $\frac{\partial^2}{\partial \theta_i \partial \theta_j}\log dF(x, \theta_0)$. Let us put

$$Y^n_{i,0} = \frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j} - E_0\left[\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j}\right]$$

$$y_{ij0} = \frac{\partial h_{i0}}{\partial \theta_j} - E_0\left[\frac{\partial h_{i0}}{\partial \theta_j}\right]$$

The characteristic function of the $Y^n_{i,}$ is

$$\varphi_n(t_{11}, \cdots, t_{hh}) = \varphi_n(t) = E_0[\exp(i\sum t_{ij}Y_{ij})]$$

$$= \left\{E_0\left[\exp\left(\frac{i}{n}\sum t_{ij}y_{ij}\right)\right]\right\}^n$$

Expanding the exponential in powers of the t's and using (vi), we find that

$$\varphi_n(t) = \left\{1 - O\left(\frac{1}{n^2}\right)\right\}^n$$

so that we have

$$\lim_{n\to\infty}\varphi_n(t) = 1$$

uniformly in every finite interval $|t_{ij}| < M$. A basic theorem on sequences of characteristic functions [4] then guarantees that for any $\epsilon > 0$

$$\lim_{n\to\infty} P_0\left\{\left|\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j} - E_0\left[\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j}\right]\right| > \epsilon\right\} = 0$$

that is to say, that $\frac{1}{n}\frac{\partial H^n_{\cdot 0}}{\partial \theta_j}$ converges stochastically to its expected value. Under the assumptions of this paragraph the same type of reasoning may be used to show that the quantities $\frac{1}{n}H^n_{\cdot 0}$, $\frac{1}{n}L^n_{\cdot 0}$, and $\frac{1}{n}\frac{\partial L^n_{\cdot 0}}{\partial \theta_j}$ all converge stochastically to their respective mean values. It will then follow from equation (3) that the functions $\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_j}$ converge stochastically to the values $E_0\left[\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_j}\right]$. In fact, it can be shown [5] that any polynomial in these functions must converge stochastically to the same polynomial in their expected values. Hence, given any $\epsilon > 0$, the probability that the determinant $\left|\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_i}\right|$ differs in samples from Π_0 from the determinant $\left|E_0\left[\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_i}\right]\right|$ by more than ϵ can be made arbitrarily small by taking n sufficiently large. Similarly, the determinant $\left|\frac{1}{n}\frac{\partial \bar{H}^n_{\cdot 0}}{\partial \theta_j}\right|$ converges stochastically to $\left|E_0\left[\frac{1}{n}\frac{\partial \bar{H}^n_{\cdot 0}}{\partial \theta_j}\right]\right|$. Thus, given any two positive numbers ϵ, ϵ', we have the relation

$$P_0\left\{\left|\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_i}\right|^+ > \left|\frac{1}{n}\frac{\partial \bar{H}^n_{\cdot 0}}{\partial \theta_i}\right|^+ - \epsilon\right\} > 1 - \epsilon'$$

(where $+$ indicates the absolute values of the determinants), provided n is sufficiently large

As in the one parameter case, the restrictions which have been put on the class of functions L and H are not entirely necessary. But it is difficult to replace them by any other set of conditions which are not obviously *ad hoc*. Let us now summarize the above results.

THEOREM 1. *If the functions $L^n_\cdot$ and $H^n_\cdot$ satisfy the conditions* (a), $\cdots$ (e), *and if*

(f) *the functions $\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_i}$ and $\frac{1}{n}\frac{\partial \bar{H}^n_{\cdot 0}}{\partial \theta_i}$ converge stochastically to their mean values;*

(g) *the large sample distribution of the functions $\frac{1}{\sqrt{n}}\bar{L}^n_{\cdot 0}$ is the same as that of the functions $\frac{1}{\sqrt{n}}\bar{H}^n_{\cdot 0}$ and is independent of the θ_0's;*

then in large samples the confidence regions derived from the $\bar{L}$'s will almost certainly be smaller than those derived from the $\bar{H}$'s, in the sense that

$$\lim_{n\to\infty} P_0\left\{\left|\frac{1}{n}\frac{\partial \bar{L}^n_{\cdot 0}}{\partial \theta_i}\right|^+ > \left|\frac{1}{n}\frac{\partial \bar{H}^n_{\cdot 0}}{\partial \theta_i}\right|^+\right\} = 1$$

unless there is linear dependence between the L's and H's.

THEOREM 2. *The assumptions of Theorem* 1 *will be satisfied if the $L_\cdot$ and $H_\cdot$ are of class K', are linearly independent, and satisfy* vi)

THEOREM 3. *For the case of only one unknown parameter, the relation*

$$\left\{E_0\left[\frac{\partial \tilde{L}_{10}^n}{\partial \theta_1}\right]\right\}^2 \geq \left\{E_0\left[\frac{\partial \tilde{H}_{10}^n}{\partial \theta_1}\right]\right\}^2$$

$\left(\text{*equality holding only in case* } H_1^n \equiv c\frac{\partial L^n}{\partial \theta_1}\right)$ *can be derived under assumptions* (a), ..., (e) *alone. Its interpretation in terms of smallest average confidence intervals depends, however, on whether or not* (g) *is satisfied.*

At first sight it may appear that the functions

$$\psi_{n1} = \frac{1}{\sqrt{n}} \sum_{j=1}^{h} b^{n1j} H_j^n$$

to which these theorems apply are too complicated to be of any practical use, involving as they do the square root of the inverse of the matrix

$$|| B_{ij}^n || = \frac{1}{n} || E[H_i^n H_j^n] ||.$$

But in employing the method of fiducial argument in the several parameter case there is no need to take the region S in the ψ space to be an interval

$$\psi_1 < \psi_1 < \bar{\psi}_1 .$$

Instead, we may take S to be the interior of the sphere

$$\sum_{i=1}^{h} \psi_i^2 < R^2 \tag{7}$$

This enables us to avoid the computation of the b^{n1j}; for

$$\sum_{i=1}^{h} \psi_{ni}^2 = \frac{1}{n} \sum_{i,j,k=1}^{h} b^{n1j} b^{n1k} H_j^n H_k^n = \frac{1}{n} \sum_{j,k=1}^{h} B^{njk} H_j^n H_k^n$$

where $|| B^{njk} ||$ is the inverse of $|| B_{jk}^n ||$.

To indicate more precisely how the function $\sum_{i=1}^{h} \psi_{n1}^2$ may be used to determine confidence regions for the parameter point θ, we note that if the distribution law of the ψ_{n1} tends to the form

$$(2\pi)^{-\frac{h}{2}} e^{-\frac{1}{2}\Sigma \psi_i^2}$$

then $\sum_{i=1}^{h} \psi_{n1}^2$, which is identically equal to $\frac{1}{n} \sum_{i,j} B^{n1j} H_i^n H_j^n$, is approximately distributed according to the χ^2 law with h degrees of freedom. We then have

$$P\left(\frac{1}{n} \sum_{i,j} B^{n1j} H_i^n H_j^n < \chi_\alpha^2\right) = \alpha \tag{8}$$

approximately, where χ_α is given by the relation

$$\frac{1}{2\Gamma(\frac{1}{2}h)} \int_0^{\chi_\alpha} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}h-1} e^{-\frac{1}{2}\chi^2} d\chi^2 = \alpha.$$

The confidence region T corresponding to a particular sample $0_n : x_1, x_2, \cdots, x_n$ consists of those points in the θ space for which $\frac{1}{n}\sum_{i,j} B^{nij} H_i^n H_j^n < \chi_\alpha^2$ when the x's are substituted in the H's. Since the region T depends on the sample, it is essentially a random variable and the probability is α that T will include the point θ_0, that is, the point in the θ-space corresponding to the values of the θ's in the population.

For example, suppose the population Π is known to have the multinomial distribution law

$$f(x_0, \cdots, x_h ; p_0, \cdots, p_h) = p_0^{x_0} \cdot \cdots \cdot p_h^{x_h}$$

In this case each x has but two possible values, 0 and 1, and

$$x_0 + \cdots + x_h = 1, \qquad p_0 + \cdots + p_h = 1. \tag{9}$$

The likelihood function for random samples 0_n drawn from Π has for its logarithm

$$L^n = \sum_{\nu=0}^{h} n_\nu \log p_\nu$$

where $n_\nu = \sum_{\alpha=1}^{n} x_{\nu\alpha}$, $x_{\nu\alpha}$ being the value of x_ν for the α-th observation. Because of (9) there are only h independent parameters, say p_i $(i = 1, \cdots, h)$. Thus

$$L_i^n = \frac{n_i}{p_i} - \frac{n_0}{p_0}$$

and

$$A_{ij}^n = \frac{\delta_{ij}}{p_i} + \frac{1}{p_0}$$

where δ_{ij} is unity if $i = j$ and 0 if $i \neq j$. It is not necessary to compute the a^{nij}, for, as we have seen,

$$\sum_{i=1}^{h} (\psi_i^n)^2 = \frac{1}{n} \sum_{i,j=1}^{n} A^{nij} L_i^n L_j^n$$

And one can immediately verify that

$$A^{nij} = \delta_{ij} p_i - p_i p_j$$

so that we have

$$\sum_{i=1}^{h} \psi_{ni}^2 = \frac{1}{n} \sum_{i,j=1}^{h} (\delta_{ij} p_i - p_i p_j) \left(\frac{n_i}{p_i} - \frac{n_0}{p_0}\right)\left(\frac{n_j}{p_j} - \frac{n_0}{p_0}\right) \tag{10}$$

Since in this case the L_i^n satisfy the conditions of the lemma, we know that $\sum_{i=1}^{h} \psi_{ni}^2$ is distributed, in large samples, approximately like χ^2 with h degrees of freedom.

As a matter of fact, (10) is precisely the Pearson χ^2 which is ordinarily used, in connection with the problem of deciding whether a sample supports the hypothesis that the population from which it has been drawn has specified values of the p's. For, making use of the fact that

$$\sum_{i=1}^{h} (n_i - np_i) + (n_0 - np_0) = 0$$

we find that

$$\frac{n_i}{p_i} - \frac{n_0}{p_0} = \sum_{j=1}^{h} A_{ij}^n (n_j - np_j)$$

so that $\sum_{i=1}^{h} \psi_{ni}^2$ reduces to

$$\frac{1}{n} \sum_{i,j=1}^{h} A_{ij}^n (n_i - np_i)(n_j - np_j) = \sum_{\nu=0}^{h} (n_\nu - np_\nu)^2 / np_\nu$$

which is the familiar form. Thus in particular the Pearson χ^2 is the best fiducial function of its type which can be formed from H's satisfying Theorem 1, in the sense that for sufficiently large samples its constituent functions $\bar{L}_i^n$ will almost certainly have a greater jacobian with respect to the parameters p_i than will the corresponding $\bar{H}_i^n$ computed from a set of H_i^n independent of the L_i^n.

The confidence regions determined by (8) when the H_i^n are replaced by the L_i^n have an associated optimum property which may be stated as

THEOREM 4: *Let Δ_0 denote the differential of $\frac{1}{n} \sum_{i,j} B^{ni,j} H_i^n H_j^n$ with respect to the θ_i, evaluated at the true parameter point θ_0. Let Δ_0^* be the corresponding differential when the H_i^n are replaced by the L_i^n. Let the H_i^n and L_i^n satisfy conditions* (i), (ii), $\cdots$, (vi) *and let the mean value of the product of two, three or four factors taken from the set $\left\{h_{i0}, \frac{\partial h_{j0}}{\partial \theta_k}\right\}$ be finite, no product containing more than two factors of the type $\frac{\partial h_{i0}}{\partial \theta_j}$. Let similar assumptions hold for the set $\left\{l_{i0}, \frac{\partial l_{i0}}{\partial \theta_i}\right\}$ where $l_{i0} = \frac{\partial \log dF_0}{\partial \theta_i}$. Then if n is sufficiently large*

$$E_0\left(\frac{1}{n}\Delta_0^{*2}\right) - E_0\left(\frac{1}{n}\Delta_0^2\right) \geq 0 \tag{11}$$

The equality in (11) *will hold for all differential vectors if and only if each h_{i0} is a linear function of the l_{i0}.*

This theorem can be proved in a straightforward manner by using the following characteristic functions

$$\varphi_H = \exp\left(i\sum_{i=1}^{h} t_i H_{i0}^n + i\sum_{i,j=1}^{h} u_{ij}\frac{\partial H_{i0}^n}{\partial\theta_j}\right) = \left[\exp\left(i\sum_{i=1}^{h} t_i h_{i0} + i\sum_{i,j=1}^{h} u_{ij}\frac{\partial h_{i0}}{\partial\theta_j}\right)\right]^n$$

$$\varphi_L = \exp\left(i\sum_{i=1}^{h} t_i L_i^n + i\sum_{i,j=1}^{h} u_{ij}\frac{\partial L_i^n}{\partial\theta_j}\right) = \left[\exp\left(i\sum_{i=1}^{h} t_i l_{i0} + i\sum_{i,j=1}^{h} u_{ij}\frac{\partial l_{i0}}{\partial\theta_j}\right)\right]^n,$$

where $u_{ij} \equiv u_{ji}$. Now

$$\Delta_0 = \frac{1}{n}\sum_{i,j,k=1}^{h}\frac{\partial B^{ij}}{\partial\theta_k}H_i^n H_j^n\,d\theta_k + \frac{2}{n}\sum_{i,j,k=1}^{h} B^{ij}\frac{\partial H_i^n}{\partial\theta_k}H_j^n\,d\theta_k$$

with a similar expression for Δ_0^*. The problem of finding the mean values $E\left(\frac{1}{n}\Delta_0^2\right)$ and $E\left(\frac{1}{n}\Delta_0^{*2}\right)$ is a matter of evaluating a set of fourth order derivatives of φ_H and φ_L at $t_i = 0$, $u_{ij} = 0$.

If the appropriate differentiations are carried out it is found that

$$E_0(\Delta_0^2) = 4n\left[\sum_{i,j,k,l} B_{kl0}C_{ki0}C_{lj0}\,d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right]$$

$$E_0(\Delta_0^{*2}) = 4n\left[\sum_{i,j,k,l} A_{ij0}\,d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right]$$

where $A_{ij0} = E_0[l_{i0}l_{j0}]$, $B_{kl0} = E_0[h_{k0}h_{l0}]$, $C_{ki0} = E_0[h_{k0}, l_{i0}]$. Denoting $E_0\left(\frac{1}{n}\Delta_0^{*2}\right) - E_0\left(\frac{1}{n}\Delta_0^2\right)$ by δ, we have

$$\delta = 4\left\{\sum_{i,j} M_{ij0}\,d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right\}$$

where $\| M_{ij0} \| = \| A_{ij0} - \sum_{k,l} B_0^{kl}C_{ki0}C_{lj0} \|$. If the h_{k0} and l_{i0} are linearly independent then $\| M_{ij0} \|$ is a positive definite matrix and hence $\sum_{i,j} M_{ij0}\,d\theta_i\,d\theta_j \equiv \delta'$ say, will be non-negative and can vanish only when all $d\theta_i$ are zero. If each h_{k0} is a linear combination of the l_{i0} and if the h_{k0} are linearly independent, then each l_{i0} is a linear combination of the h_{k0}. In this case it can be readily shown that every element in $\| M_{ij0} \|$ will vanish, and hence $\delta' \equiv 0$.

In case some of the h_{i0} are linearly dependent on the l_{j0}, it can be shown that δ' is positive semidefinite, that is, there exists no differential vector for which δ' is negative, although there will exist non-zero differential vectors for which δ' is zero.

It can be shown under the assumptions made in Theorem 4 that $\frac{1}{n}(\Delta_0^{*2} - \Delta_0^2)$ actually converges stochastically to $4\delta'$, and thus if the h_{i0} and l_{i0} are linearly independent, the difference $\frac{1}{n}(\Delta_0^{*2} - \Delta_0^2)$ converges stochastically to a positive number. Stated in another way: for sufficiently large samples, the square of the differential change in $\frac{1}{n}\sum_{i,j} A^{ij} L_i^n L_j^n$, for a given change $d\theta_i$ in the θ_i from the values θ_{i0}, will almost certainly exceed that of $\frac{1}{n}\sum_{i,j} B^{ij} H_i^n H_j^n$. The statistical interpretation of this result amounts to the following: by taking sufficiently large samples, we can make it as certain as we please that the confidence regions for locating θ_0 determined by using $\frac{1}{n}\sum_{i,j} A^{ij} L_i^n L_j^n$ in (8) will be smaller than those determined by using $\frac{1}{n}\sum_{i,j} B^{ij} H_i^n H_j^n$ in (8).

REFERENCES

[1] S. S. Wilks, "Shortest Average Confidence Intervals from Large Samples," *Annals of Mathematical Statistics*, vol. IX, No 3 (1938), pp. 166–175.

[2] S. S Wilks, "Optimum Fiducial Regions for Simultaneous Estimation of Several Population Parameters from Large Samples," (Abstract), *Annals of Mathematical Statistics*, vol. X, No. 1 (1939), p. 85

[3] C. C. MacDuffee, *The Theory of Matrices*, Ergebnisse Series, vol. II, No. 5, Berlin, Julius Springer (1933), p. 87.

[4] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, Cambridge University Press, 1937.

[5] J. L Doob, "Limiting Distributions of Certain Statistics," *Annals of Mathematical Statistics*, vol. VI, No 3 (1935), p 164.

PRINCETON UNIVERSITY, PRINCETON, N. J.

ON SOME PROPERTIES OF MULTIDIMENSIONAL DISTRIBUTIONS

By J. Lukomski

If, in a system of random variables $x_1, x_2, \cdots, x_n$, some variables are connected by a functional (exact) dependence, the n-dimensional distribution law has a degenerated character. In other words, in this case the probability is not distributed over the whole n-dimensional space, but is concentrated on a manifold of a smaller number of dimensions which may be called the *skeleton of the distribution.*

The character and the dimensionality of this manifold are determined by the character and the number of functional connections between the variables $x_1, x_2, \cdots, x_n$. If all these connections are linear, the skeleton will be a linear manifold (hyperplane). The investigation of the skeleton of distribution represents obviously an interest from the theoretical as well as from the practical point of view.

In the present paper we establish some criteria which enable us to determine, for any distribution possessing finite moments of the first and second order, the linear skeleton and to find the variations of the dimensionality of this manifold when the variables are subjected to a linear transformation.[1]

We also apply the obtained results to the case of a multidimensional normal distribution (generalized by H. Cramér to the case of linear dependence between variables).

§1

Let

$$x_1, x_2, \cdots, x_n \tag{1}$$

be a system of random variables defined in the n-dimensional euclidean space R_n by the multidimensional distribution function $F(x_1, x_2, \cdots, x_n)$. The function F is defined on all Borel sets in R_n. We assume the existence of the following moments:

$$E(x_i) = \int\int \cdots \int_{R_n} x_i \, dd \cdots dF(x_1, x_2, \cdots x_n) = 0$$

$$E(x_i x_j) = \int\int \cdots \int_{R_n} x_i x_j \, dd \cdots dF(x_1, x_2, \cdots x_n) = \mu_{ij}$$

where the integrals are to be understood in the sense of Lebesgue-Radon

[1] The questions of degeneracy of a statistical distribution were for the first time considered—from a somewhat different point of view—by R. Frisch [1].

If the variables $x_1, x_2, \cdots, x_n$ are connected by a relation of the form $C_1x_1 + C_2x_2 + \cdots + C_nx_n = 0$ $(\Sigma C^2 \neq 0)$ (are linearly dependent), we call this relation a *linear bond of the distribution* F.

We shall call a system of linear bond of the distribution F *complete*, if all bonds of the system are linearly independent and every linear bond of the distribution depends linearly on the bonds of the system.

By the (linear) *decrement* of the distribution F (we denote it by $k(F)$ or simply k) we understand the number of bonds in a complete system. We may, correspondingly, call the difference between the number of variables and the decrement of the distribution the (linear) *rank* of the distribution, or the dimensionality of the linear skeleton.

The decrement (rank) is given by the following

THEOREM 1.[2] *The decrement (rank) of the distribution F is equal to the decrement*[3] *(rank) of the matrix*

$$\| \mu_{ij} \| \qquad i, j = 1, 2, \cdots n$$

of the moments of the second order of this distribution; that is

$$k(F) = k(\| \mu_{ij} \|), \qquad i, j = 1, 2, \cdots n \tag{2}$$

PROOF. Consider the form

$$v = t_1x_1 + t_2x_2 + \cdots + t_nx_n \tag{3}$$

where $t_1, t_2, \cdots, t_n$ are arbitrary real numbers, not all equal to zero. Let

$$Q^2 = E(v^2) = \int\int \cdots \int_{R_n} (t_1x_1 + t_2x_2 + \cdots + t_nx_n)^2 \, dd \cdots$$

$$\cdots dF(x_1, x_2, \cdots x_n) \tag{4}$$

$$= \sum_{i,j=1}^{n} t_it_j \int\int \cdots \int_{R_n} x_ix_j \, dd \cdots dF(x_1, x_2, \cdots x_n) = \sum_{i,j=1}^{n} t_it_j\mu_{ij}$$

Q^2 is a non-negative quadratic form in the variables $t_1, t_2, \cdots, t_n$. The system of values $t_1, t_2, \cdots, t_n$, for which the expression (3) becomes zero is a double point of the form Q^2.

The coordinates of the double point can be found from the system of homogeneous equations:

$$\begin{aligned} &\mu_{11}t_1 + \mu_{12}t_2 + \cdots + \mu_{1n}t_n = 0 \\ &\mu_{21}t_1 + \mu_{22}t_2 + \cdots + \mu_{2n}t_n = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\mu_{n1}t_1 + \mu_{n2}t_2 + \cdots + \mu_{nn}t_n = 0. \end{aligned} \tag{5}$$

[2] This theorem was proved by a different method by R. Frisch [1].

[3] By the decrement of a (rectangular) matrix we call, after B. Kagan, the difference between the number of its rows and its rank.

It is, however, known that the number of the independent double points of the form, Q^2, i.e the number of linearly independent untrivial solutions of the system (5) is equal to the decrement of the matrix $|| \mu_{ij} ||$, $i, j = 1, 2, \cdots n$.

Consequently, there exist only $k(|| \mu_{ij} ||)$ independent linear connections between the variables $x_1, x_2, \cdots, x_n$, which proves the theorem.

Hence it follows that the variables $x_1, x_2, \cdots, x_n$ are linearly independent ($k(F) = 0$) if and only if the form Q^2 is positive definite and, consequently, the discriminant $| \mu_{ij} |$ of the form is positive.

The following two theorems may be used for determination of a complete system of linear bonds. The first of them is a special case of the second, but is stated separately in order to simplify the proof.

THEOREM 2. *If $k(F) = 1$, we obtain the linear bond of the distribution by replacing in the determinant on the left hand side of the equation*

$$\begin{vmatrix} \mu_{11} & \mu_{12} \cdots \mu_{1n} \\ \mu_{21} & \mu_{22} \cdots \mu_{2n} \\ \cdots\cdots\cdots\cdots \\ \mu_{n1} & \mu_{n2} \cdots \mu_{nn} \end{vmatrix} = 0 \tag{6}$$

the elements of one (arbitrary) row by $x_1, x_2, \cdots, x_n$ respectively.

For instance, replacing the first row, we have

$$\begin{vmatrix} x_1 & x_2 \cdots x_n \\ \mu_{21} & \mu_{22} \cdots \mu_{2n} \\ \cdots\cdots\cdots\cdots \\ \mu_{n1} & \mu_{n2} \cdots \mu_{nn} \end{vmatrix} = 0. \tag{7}$$

PROOF. Since the decrement of the matrix $|| \mu_{ij} ||$, $i, j = 1, 2, \cdots n$ is equal to 1, for the unique nontrivial independent solution of the system (5) ($t_1, t_2, \cdot \ , t_n$) may be taken, as we know, the system of algebraical supplements of the elements of any row of the determinant $| \mu_{ij} |$, $i, j = 1, 2, \cdots n$. (Among the algebraical supplements of elements of each row there is at least one different from zero, since the algebraical supplements of corresponding elements of any pair of rows are proportional to each other.)

Hence, since $t_1x_1 + t_2x_2 + \cdots + t_nx_n = 0$, the theorem follows.

THEOREM 3. *If $k(F) > 0$, we obtain a complete system of linear bonds of the distribution F replacing in each of the k equations*

$$\begin{vmatrix} \mu_{k1} & \mu_{k,k+1} \cdots \mu_{kn} \\ \mu_{k+1,1} & \mu_{k+1,k+1} \cdots \mu_{k+1,n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \mu_{n1} & \mu_{n,k+1} \cdots \mu_{nn} \end{vmatrix} = 0, \qquad i = 1, 2, \cdots k \tag{8}$$

one (arbitrary) row of the determinant respectively by $x_i, x_{k+1}, \cdots, x_n$, where $x_{k+1}, \cdots, x_n$ are chosen in such a way that

$$\begin{vmatrix} \mu_{k+1,k+1} \cdots \mu_{k+1,n} \\ \cdots\cdots\cdots\cdots \\ \mu_{n,k+1} \cdots \mu_{nn} \end{vmatrix} > 0.$$

Replacing, for example, the first rows, we obtain:

$$(9)\qquad \begin{vmatrix} x_1 & x_{k+1} & \cdots & x_n \\ \mu_{k+1,1} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0$$

$$\begin{vmatrix} x_2 & x_{k+1} & \cdots & x_n \\ \mu_{k+1,2} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n2} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\begin{vmatrix} x_k & x_{k+1} & \cdots & x_n \\ \mu_{k+1,k} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{nk} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0.$$

PROOF. The theorem is already proved for $k(F) = 1$ (Theorem 2) We have to prove it for $k(F) > 1$.

Let us in the first place show that the matrix $|| \mu_{ij} ||$, $i, j = 1, 2, \cdots n$ possesses at least one positive chief algebraical supplement of the order $n - k$.

In fact, in the system of n variables $x_1, x_2, \cdots, x_n$, connected by k independent linear relations there must exist a subsystem of $n - k$ linearly independent variables. Let these variables be $x_{k+1}, x_{k+2}, \cdots, x_n$. The determinant of the moments of the second order of this subsystem: $| \mu_{ij} |$, $i, j = k+1, \cdots, n$ is different from zero and, by the property of Gramm's determinants, is positive. Further, each of the subsystems $x_i, x_{k+1}, \cdots, x_n$, is subjected to the distribution law $F_i(x_i, x_{k+1}, \cdots, x_n)$ with the decrement $k_i = 1$ and, consequently, by Theorem 2, the relations (9) are satisfied. (Arguing as before we find that any (not necessarily the first) row in each of the determinants in (8) may be replaced by $x_i, x_{k+1}, \cdots x_n$).

In order to show the independence of the relations (9), write the system (9) in the form:

$$(9')\qquad \sum_{j=1}^{n} C_{ij} x_j = 0, \qquad i = 1, 2, \cdots, k$$

and consider the matrix of its coefficients:

$$(10)\qquad \left\| \begin{matrix} C_{11} & 0 & \cdots & 0 & C_{1,k+1} & C_{1,k+2} & \cdots & C_{1n} \\ 0 & C_{22} & \cdots & 0 & C_{2,k+1} & C_{2,k+2} & \cdots & C_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_{kk} & C_{k,k+1} & C_{k,k+2} & \cdots & C_{kn} \end{matrix} \right\|.$$

The matrices (10) have the rank k, since the determinant of order k

$$\begin{vmatrix} C_{11} & 0 & \cdots & 0 \\ 0 & C_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_{kk} \end{vmatrix} = C_{11} \cdot C_{22} \cdots C_{kk}$$

belonging to the matrix, is positive; this follows from

$$C_{11} = C_{22} = \cdots = C_{kk} = |\mu_{ij}| > 0, \qquad i, j = k+1, \cdots, n.$$

Thus the independence of the relations (9) is proved and the theorem is established.

§2

In this section we consider the question of the variation of decrement of the distribution in the case when the variables are subjected to a linear transformation.

Let $x_1, x_2, \cdots, x_n$ be a system (1) of random variables and

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned} \tag{11}$$

a system of linear forms in the variables (1).

The distribution function of the variables $u_1, u_2, \cdots, u_m$ we denote by F_1, the decrement of the distribution by $k(F_1)$, or, shorter, by k_1.

The two systems of equations (11) and (9) form together the system:

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \\ 0 &= a_{m+1,1}x_1 + a_{m+1,2}x_2 + \cdots + a_{m+1,n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 &= a_{m+k,1}x_1 + a_{m+k,2}x_2 + \cdots + a_{m+k,n}x_n \end{aligned} \tag{12}$$

where the last k equations represent, in new notation, the equations (9).

We call the matrix of the coefficients of the variables in the system (12): $\| a_{ij} \|$ $i = 1, 2, \cdots, m+k$; $j = 1, 2, \cdots, n$, *the elongated matrix of the transformation.*

We prove the following

THEOREM 4. *The decrement of the distribution $F_1(u_1, u_2, \cdots u_m)$ is equal to the decrement of the elongated matrix of the transformation.*

$$k(F_1) = k(\| a_{ij} \|). \qquad \begin{matrix} i = 1, 2, \cdots, m+k \\ j = 1, 2, \cdots, n \end{matrix} \tag{13}$$

PROOF. Consider a system of forms in arbitrary linearly independent parameters $\xi_1, \xi_2, \cdots, \xi_n$:

$$
\begin{aligned}
v_1 &= a_{11}\xi_1 + a_{12}\xi_2 + \cdots + a_{1n}\xi_n \\
v_2 &= a_{21}\xi_1 + a_{22}\xi_2 + \cdots + a_{2n}\xi_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_m &= a_{m1}\xi_1 + a_{m2}\xi_2 + \cdots + a_{mn}\xi_n \\
v_{m+1} &= a_{m+1,1}\xi_1 + a_{m+1,2}\xi_2 + \cdots + a_{m+1,n}\xi_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_{m+k} &= a_{m+k,1}\xi_1 + a_{m+k,2}\xi_2 + \cdots + a_{m+k,n}\xi_n
\end{aligned} \tag{14}
$$

such that the matrix of the system (14) coincides with the elongated matrix of the transformation.

For

$$
v_{m+1} = 0, \quad v_{m+2} = 0, \cdots, \quad v_{m+k} = 0 \tag{15}
$$

the system (14) reduces to the system (12).

If the decrement of the matrix of the system is equal to s, there exist only $m + k - s$ linearly independent forms v_i, and each of the remaining s forms is a linear combination of the first.

By Steinitz's theorem we can always include in a subsystem of independent forms the forms $v_{m+1}, \cdots, v_{m+k}$ (since these forms are independent).

Denoting all forms of the subsystem by $v_{s+1}, \cdots, v_m, v_{m+1}, \cdots, v_{m+k}$, let us write the s relations connecting each of the remaining forms with the forms of our subsystem in the form:

$$
\begin{aligned}
&g_{11}v_1 + g_{1,s+1}v_{s+1} + \cdots + g_{1m}v_m + g_{1,m+1}v_{m+1} + \cdots + g_{1,m+k}v_{m+k} = 0 \\
&g_{22}v_2 + g_{2,s+1}v_{s+1} + \cdots + g_{2m}v_m + g_{2,m+1}v_{m+1} + \cdots + g_{2,m+k}v_{m+k} = 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&g_{ss}v_s + g_{s,s+1}v_{s+1} + \cdots + g_{sm}v_m + g_{s,m+1}v_{m+1} + \cdots + g_{s,m+k}v_{m+k} = 0
\end{aligned} \tag{16}
$$

where $g_{11}, g_{22}, \cdots, g_{ss} \neq 0$.

Assigning to the variables in these equations the values (15) we clearly obtain s linear relations between the variables $u_1, u_2, \cdots, u_m$

$$
\begin{aligned}
g_{11}u_1 + g_{1,s+1}u_{s+1} + \cdots + g_{1m}u_m &= 0 \\
g_{22}u_2 + g_{2,s+1}u_{s+1} + \cdots + g_{2m}u_m &= 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
g_{ss}u_s + g_{s,s+1}u_{s+1} + \cdots + g_{sm}u_m &= 0.
\end{aligned} \tag{17}
$$

The equations (17) are linearly independent, since the matrix of the system (17)

$$\begin{Vmatrix} g_{11} & 0 \cdots 0 & g_{1,s+1} \cdots g_{1m} \\ 0 & g_{22} \cdots 0 & g_{2,s+1} \cdots g_{2m} \\ \cdots\cdots\cdots\cdots & \cdots\cdots & \\ 0 & 0 \cdots g_{ss} & g_{s,s+1} \cdots g_{sm} \end{Vmatrix}$$

has the rank s; this follows from the fact that it contains the determinant

$$\begin{vmatrix} g_{11} & 0 \cdots 0 \\ 0 & g_{22} \cdots 0 \\ \cdots\cdots\cdots & \\ 0 & 0 \cdots g_{ss} \end{vmatrix} = g_{11} \cdot g_{22} \cdots g_{ss}$$

of the order s, which is different from zero.

We proceed now to prove that there exists no other linear relation between the variables u_i linearly independent of the relations (17)

From the equations (17) the variables $u_1, u_2, \cdots, u_s$ may be determined as linear combinations of the variables $u_{s+1}, \cdots, u_m$ (we suppose that $m > s$, since for $m = s$ the proposition under consideration is trivial).

It is thus to be proved that the variables $u_{s+1}, \cdots, u_m$ are linearly independent (since every new linear relation between the variables $u_1, u_2, \cdots, u_m$, independent of (17) must, after corresponding substitutions, lead to a linear relation between $u_{s+1}, \cdots, u_m$).

In the equations (12) the linearly independent variables $u_{s+1}, \cdots, u_{m+k}$ are linear forms in n linearly independent parameters $\xi_1, \xi_2, \cdots, \xi_n$.

We may instead of the $\xi_1, \xi_2, \cdots, \xi_n$ take for the system of linearly independent parameters $v_{m+1}, \cdots, v_{m+k}, \xi_{k+1}, \cdots, \xi_n$ (changing the indices of the ξ in an appropriate manner), defining $\xi_1, \xi_2, \cdots, \xi_k$[4] by the system of equations

$$\begin{aligned} v_{m+1} &= a_{m+1,1}\xi_1 + \cdots + a_{m+1,n}\xi_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ v_{m+k} &= a_{m+k,1}\xi_1 + \cdots + a_{m+k,n}\xi_n \end{aligned}$$

which is always possible, since the forms $v_{m+1}, \cdots, v_{m+k}$ are independent.

Substituting the expressions obtained for the $\xi_1, \xi_2, \cdots, \xi_n$ into the forms $v_{s+1}, \cdots, v_m$, we find

$$\begin{aligned} v_{s+1} &= \varphi_{s+1}(v_{m+1}, \cdots v_{m+k}) + \psi_{s+1}(\xi_{k+1}, \cdots \xi_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ v_m &= \varphi_m(v_{m+1}, \cdots v_{m+k}) + \psi_m(\xi_{k+1}, \cdots \xi_n) \end{aligned} \tag{18}$$

where φ and ψ are linear forms in the corresponding arguments.

The variables $v_{s+1}, \cdots, v_m$ remain, of course, independent.

---

[4] The indices of the ξ adequately chosen

Performing in the equations (18) the substitution (15), we obtain:

$$\begin{aligned} u_{s+1} &= \psi_{s+1}(\xi_{k+1}, \cdots \xi_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= \psi_m(\xi_{k+1}, \cdots \xi_n). \end{aligned} \tag{19}$$

If there exists a linear dependence between the $u_{s+1}, \cdots, u_m$, we can find $\alpha_{s+1}, \cdots, \alpha_m$, not all equal to zero, such that

$$\alpha_{s+1}u_{s+1} + \cdots + \alpha_m u_m = 0. \tag{20}$$

Multiplying the equations (18) by the coefficients $\alpha_{s+1}, \cdots, \alpha_m$ respectively, and adding, we obtain, by virtue of (19) and (20)

$$\alpha_{s+1}v_{s+1} + \cdots + \alpha_m v_m = \alpha_{s+1}\varphi_{s+1}(v_{m+1}, \cdots v_{m+k}) + \cdots + \alpha_m\varphi_m(v_{m+1}, \cdots v_{m+k})$$

i.e. the variables $v_{s+1}, \cdots, v_{m+k}$ are linearly dependent, which contradicts the assumption.

The required proposition is thus proved.

It follows that the s equations (17) form a complete system of bonds of the distribution F_1, which proves our theorem.

The moments of the second order of the distribution F_1 are connected with the moments of the distribution F by the following formulae

$$\begin{aligned} v_{ij} &= E(u_i u_j) = E\left[\left(\sum_{r=1}^{n} a_{ir}x_r\right)\left(\sum_{s=1}^{n} a_{js}x_s\right)\right] \\ &= \sum_{r,s=1}^{n} a_{ir}a_{js}E(x_r x_s) = \sum_{r,s=1}^{n} a_{ir}a_{js}\mu_{rs} \qquad (i, j = 1, 2, \cdots m). \end{aligned} \tag{21}$$

§3

Let the normal law of distribution G (generalized by H. Cramér) be given by its multidimensional characteristic function [2], [3]:

$$\begin{aligned} f(t_1, t_2, \cdots t_n) &= \int\int \cdots \int_{R_n} e^{i(t_1x_1 + t_2x_2 + \cdots + t_nx_n)}\, d\,d \cdots dG(x_1, x_2, \cdots x_n) \\ &= e^{-\frac{1}{2}Q^2} \end{aligned} \tag{22}$$

where $Q^2 = \sum_{r,s=1}^{n} c_{rs}t_r t_s$ $(c_{rs} = c_{sr})$ is a non-negative quadratic form in the real variables $t_1, t_2, \cdots, t_n$. (The integrals, as above, to be understood in the sense of Lebesgue-Radon.)

As is easily seen, the coefficients c_{rs} are the moments of the second order of the distribution G for which

$$\mu_{rs} \equiv E(x_r x_s) \equiv i^2\left[\frac{\partial^2 f}{\partial t_r\, \partial t_s}\right]_{\Sigma t^2 = 0} = c_{rs}.$$

If Q^2 is positive definite, we have a proper normal distribution.

If Q^2 is non-negative, the distribution G possesses a positive decrement.

The decrement and the linear bonds of the distribution may be determined from the matrix of the coefficients $\| c_{rs} \|$ $r, s = 1, 2, \cdots, n$ on ground of the general theorems of §1.

Let, as before,

$$
\begin{aligned}
u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{aligned} \tag{11}
$$

be a system of linear forms in the variables $x_1, x_2, \cdots, x_n$. We shall prove the following

THEOREM 5. *The variables $u_1, u_2, \cdots, u_m$ are subject to the generalized normal distribution law the decrement of which is equal to the decrement of the elongated matrix of the transformation*

$$
\left\| \begin{array}{cccc}
a_{11} & a_{12} & \cdots & a_{1n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} \\
a_{m+1,1} & a_{m+1,2} & \cdots & a_{m+1,n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{m+k,1} & a_{m+k,2} & \cdots & a_{m+k,n}
\end{array} \right\|.
$$

PROOF. Consider the characteristic function of the distribution $G_1(u_1, u_2, \cdots u_m)$,

$$
f_1(t_1, t_2, \cdots t_m) = \int\int \cdots \int_{R_m} e^{i(t_1u_1 + t_2u_2 + \cdots + t_mu_m)}\, d\,d \cdots dG_1(u_1, u_2 \cdots, u_m). \tag{23}
$$

Performing in this expression the substitution (11), we obtain

$$
\begin{aligned}
& f_1(t_1, t_2, \cdots t_m) \\
&= \int\int \cdots \int_{R_n} e^{i\left(t_1 \sum_{j=1}^{n} a_{1j}x_j + t_2 \sum_{j=1}^{n} a_{2j}x_j + \cdots + t_m \sum_{j=1}^{n} a_{mj}x_j\right)}\, d\,d \cdots \\
&\qquad \cdots dG_1\left\{\sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{2j}x_j, \cdots \sum_{j=1}^{n} a_{mj}x_j\right\} \\
&= \int\int \cdots \int_{R_n} e^{i\left(x_1 \sum_{p=1}^{m} a_{p1}t_p + x_2 \sum_{p=1}^{m} a_{p2}t_p + \cdots + x_n \sum_{p=1}^{m} a_{pn}t_p\right)}\, d\,d \\
&\qquad \cdots dG(x_1, x_2, \cdots x_n).
\end{aligned} \tag{24}
$$

($d\,d \cdots dG(x_1, x_2, \cdots x_n)$ in the expression (24) does not, in general, coincide with $d\,d \cdots dG(x_1, x_2, \cdots x_n)$ in the expression (22)).

Taking into account (22), we obtain

$$
f_1 = e^{-\frac{1}{2}Q_1}
$$

where

$$Q_1^2 = \sum_{r,s=1}^{n} \left\{ c_{rs} \left(\sum_{p=1}^{m} a_{pr} t_p \right) \left(\sum_{q=1}^{m} a_{qs} t_q \right) \right\}$$

$$= \sum_{r,s=1}^{n} \left\{ c_{rs} \sum_{p,q=1}^{m} a_{pr} a_{qs} t_p t_q \right\} \tag{26}$$

$$= \sum_{p,q=1}^{m} \left\{ t_p t_q \sum_{r,s=1}^{n} a_{pr} a_{qs} c_{rs} \right\} = \sum_{p,q=1}^{m} t_p t_q v_{pq}.$$

Q_1^2 is a non-negative quadratic form in $t_1, t_2, \cdots, t_m$, the coefficients of which coincide with the moments of the second order of the distribution $G_1(u_1, u_2, \cdots u_m)$.

Consequently, the distribution G_1 is a generalized normal distribution.

By Theorem 4 the decrement of the distribution G_1 is equal to the decrement of the matrix $|| a_{pr} ||$ $p = 1, 2, \cdots, m + k$; $r = 1, 2, \cdots, n$, the last k rows of which consist of the coefficients of the complete system of linear bonds of the distribution G.

Let now $x_1, x_2, \cdots, x_n$ be a system of random variables subjected to a proper Gaussian law. The density function of the distribution of the system is

$$y = Ce^{-\chi^2} = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} \sqrt{\mu_{ii}} \sqrt{R}} \exp\left[-\tfrac{1}{2} R \left\{ \sum_{i=1}^{n} R_{ii} \frac{x_i^2}{\mu_{ii}} + 2 \sum_{i} \sum_{j>i} R_{ij} \frac{x_i x_j}{\sqrt{\mu_{ii} \mu_{jj}}} \right\} \right]$$

where

$$R = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{2n} & \cdots & 1 \end{vmatrix}.$$

R_{ij} are the algebraical supplements in R, $r_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{ii} \mu_{jj}}}$, and χ^2 is a positive definite quadratic form in the variables $x_1, x_2, \cdots, x_n$.

Again let $u_1, u_2, \cdots, u_m$ be a system of linear forms in the variables $x_1, x_2, \cdots, x_n$

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n. \end{aligned} \tag{11}$$

Then from Theorem 5 follows the

COROLLARY. *The random variables* $u_1, u_2, \cdots, u_m$ *are subject to the* m*-dimensional properly normal distribution law of Gauss if and only if the matrix*

$$\begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Vmatrix}$$

of the system of forms (11) *has the rank* m.

REFERENCES

[1] R. Frisch, "Correlation and Scatter in Statistical Variables," *Nordsk Statistisk Tidskrift*, vol. 1(1929), 36–102.

[2] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeitschr.*, vol. 41(1936),

[3] H. Cramér, *Random Variables and Probability Distributions*, Cambridge 1937

Moscow, U. S. S. R.

ON A CLASS OF DISTRIBUTIONS THAT APPROACH THE NORMAL DISTRIBUTION FUNCTION[1]

By George B. Dantzig

1. **Formulation of the Problem.** An important property of a sequence of binomial coefficients is that, when suitably normalized and transformed, it converges to the normal distribution.[2] The object of this paper is to exhibit a large class of other sequences which also possess this property.

The Pascal recurrence formula may be taken as the defining property of the binomial coefficients. Let the combination of n things taken x at a time be denoted by $\binom{n}{x}$. If we set $f_n(x) = (\frac{1}{2})^n \cdot \binom{n}{x}$ for $0 \leq x \leq n$ and $f_n(x) = 0$ for $x < 0$ or $x > n$, then $f_n(x)$ is defined for all integers x. With this notation Pascal's recurrence formula, $\binom{n}{x} = \binom{n-1}{x} + \binom{n-1}{x-1}$, may be written

$$f_n(x) = \tfrac{1}{2}\,[f_{n-1}(x) + f_{n-1}(x-1)], \tag{1}$$

where this new form is valid for all integers x extending from $-\infty$ to $+\infty$.

In order to generalize, we may consider a sequence of distributions $f_1(x)$, $f_2(x), \cdots, f_n(x), \cdots$ each defined in terms of the preceding one by means of the recurrence formula

$$f_n(x) = \frac{1}{a_n+1}\,[f_{n-1}(x-0) + f_{n-1}(x-1) + f_{n-1}(x-2) + \cdots + f_{n-1}(x-a_n)], \tag{2}$$

where the x are integers, and a_n is a positive integer which may change in value from one distribution to the next. The problem is to find conditions under which $f_n(x)$, in normalized form, approaches the normal distribution. The normalization of $f_n(x)$ is effected by the affine transformation

$$u = \frac{x - \bar{x}_n}{\sigma_n}; \qquad \varphi_n(u) = f_n(x), \tag{3}$$

---

[1] Presented November 21, 1938 before a joint meeting of the Columbia Mathematics Club and the Statistical Seminar of the Graduate School of the Department of Agriculture; also December 10, 1938 before a meeting of the American Mathematical Association at the University of Maryland

[2] Due to DeMoivre, 1731 By a variable distribution approaching the normal distribution, we mean that the integral under the variable distribution between any two limits approaches the corresponding integral under the normal curve

where $\bar{x}_n$ and σ_n are the mean and standard deviation of the distribution $f_n(x)$. The normal (cumulative) distribution function is taken in the standard form

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{1}{2}x^2}\, dx\,. \tag{4}$$

The theorem whose proof forms the theme of this paper may be stated as follows:

THEOREM: *A necessary and sufficient condition that* $\varphi_n(u) \to \varphi(u)$ *as* $n \to \infty$ *is that* $\Gamma = 0$, *where*

$$\Gamma = \operatorname*{Lim}_{n\to\infty} \sum_{i=2}^{n} \gamma_i^2 \Big/ \left(\sum_{i=2}^{n} \gamma_i\right)^2; \qquad 4\gamma_i = a_i^2 + 2a_i\,. \tag{5}$$

2. Liapounoff Condition; the general case. The recurrence formula (2) is a special case of the most general linear recurrence formula

$$f_n(x) = \sum_{i=-\infty}^{+\infty} g_n(i) f_{n-1}(x - i), \tag{6}$$

where $g_n(i)$ are a given set of weight functions generating the sequence $f_1(x)$, $f_2(x), \cdots, f_n(x), \cdots$. We may form the recurrence formula (2) by setting

$$\begin{aligned} g_n(i) &= \frac{1}{a_n + 1} \qquad \text{if} \qquad 0 \le i \le a_n\,, \\ g_n(i) &= 0 \qquad\quad\ \ \text{if} \qquad i < 0 \text{ or } i > a_n\,. \end{aligned} \tag{7}$$

Let $F_k(t) = \sum_{x<t} f_k(x)$ express[3] the probability that a variable $x_k < t$, where the distribution function of x_k is defined as $f_k(x)$; and in a similar manner let the probability that a variable $s_k < t$ be given by $G_k(t) = \sum_{x<t} g_k(x)$ By summing $f_n(x)$ for all x less than t, we obtain

$$F_n(t) = \sum_{i=-\infty}^{+\infty} F_{n-1}(t - i) g_n(i) = \int_{-\infty}^{+\infty} F_{n-1}(t - i)\, dG_n(i)\,, \tag{8}$$

where we have replaced the summation by a Stieltjes Integral. In the latter form the integral gives, in general, the probability that the *sum* of two independent variables x_{n-1} and s_n is less than t From the above equation we see that the probability that $x_{n-1} + s_n < t$ is the same as that of $x_n < t$, so that we may set $x_n = x_{n-1} + s_n$. By iteration one obtains

$$x_n = s_1 + s_2 + \cdots + s_n \tag{9}$$

for all n. Thus we have established that *if a distribution function of a variable* s_k *is defined as* $g_k(x)$, *then the distribution function of the sum* $s_1 + s_2 + \cdots + s_n = x_n$ *is* $f_n(x)$.

---

[3] The summation extends over all values x less than t.

The limit of the distribution function of the sum of n independent variables as $n \to \infty$ has been considered by Laplace, Liapounoff, Lindeberg, and others. We shall make use of a sufficient condition given by Liapounoff that the normalized distribution function of x_n approaches $\varphi(u)$.

LAPLACE-LIAPOUNOFF THEOREM:[4] *A sufficient condition for the normalized distribution function of the sum of n independent variables $s_1, s_2, \cdots, s_n$ to approach the normal distribution function with increasing n is $\Gamma' = 0$, where*

$$\Gamma' = \operatorname*{Lim}_{n\to\infty} \frac{M_4(1) + M_4(2) + \cdots + M_4(n)}{[M_2(1) + M_2(2) + \cdots + M_2(n)]^2}, \tag{10}$$

and where $M_2(k)$ and $M_4(k)$ are defined as the second and fourth moments of s_k whose distribution is $g_k(x)$.

Thus we have shown that if a sequence of distributions $f_n(x)$ is defined by the general linear recurrence formula (6),

$$f_n(x) = \sum_{i=-\infty}^{+\infty} g_n(i) \cdot f_{n-1}(x - i),$$

then a sufficient condition that $\varphi_n(u) \to \varphi(u)$ as $n \to \infty$ is given by $\Gamma' = 0$, where $\varphi_n(u)$ is the normalized form of $f_n(u)$.

3. Sufficiency of the Condition $\Gamma = 0$. We may simplify the condition $\Gamma' = 0$ for the more restricted case of a sequence of distributions defined by the recurrence formula (2). In general, the second and fourth moments of $g_n(x)$ are given by

$$\begin{aligned} M_2(k) &= \sum_{x=-\infty}^{+\infty} g_k(x)(x - \bar{s}_k)^2, \\ M_4(k) &= \sum_{x=-\infty}^{+\infty} g_k(x)(x - \bar{s}_k)^4, \end{aligned} \tag{11}$$

where $\bar{s}_k$ is the mean value of the distribution. Equations (7) give the special values of $f_k(x)$; substituting these values in (11), and remembering the Bernoulli summation by which $1^p + 2^p + 3^p + \cdots + n^p$ may be expressed as a polynomial in n of degree $p + 1$, we obtain

$$\begin{aligned} M_2(k) &= \sum_{x=0}^{a_k} \frac{1}{a_k + 1} \left(x - \frac{1}{2} a_k\right)^2 = \frac{1}{3}\left[\frac{a_k^2 + 2a_k}{4}\right] = \frac{1}{3}\gamma_k, \\ M_4(k) &= \sum_{x=0}^{a_k} \frac{1}{a_k + 1} \left(x - \frac{1}{2} a_k\right)^4 \\ &= \frac{1}{5}\left[\frac{a_k^2 + 2a_k}{4}\right]^2 - \frac{1}{15}\left[\frac{a_k^2 + 2a_k}{4}\right] = \frac{1}{5}\gamma_k^2 - \frac{1}{15}\gamma_k; \end{aligned} \tag{12}$$

[4] J. V. Uspensky, *Introduction to Mathematical Probability* (McGraw-Hill, 1937), pages 284-292; the theorem is proved there by the method of characteristic functions.

whence by substitution in (10), Γ' becomes

$$\Gamma' = \operatorname*{Lim}_{n\to\infty} \frac{\frac{1}{5}\sum_{i=2}^{n}\gamma_i^2 - \frac{1}{15}\sum_{i=2}^{n}\gamma_i + M_4(1)}{\left[\frac{1}{3}\sum_{i=2}^{n}\gamma_i + M_2(1)\right]^2}. \tag{13}$$

Since $a_i \geq 1$, $\gamma_i \geq 3/4$, and thus $\sum_{i=2}^{n}\gamma_i \to \infty$ as $n \to \infty$, we may reduce Γ' in the limit to

$$\Gamma' = \frac{3}{5}\operatorname*{Lim}_{n\to\infty} \sum_{i=2}^{n}\gamma_i^2 \Big/ \left[\sum_{i=2}^{n}\gamma_i\right]^2. \tag{14}$$

Since $\Gamma' = \frac{3}{5}\Gamma$, the Liapounoff condition $\Gamma' = 0$ for normality becomes by (5), $\Gamma = 0$.

4. **Necessity of the Condition $\Gamma = 0$.** A necessary condition for normality can be found by noting that if $\varphi_n(u)$ approaches $\varphi(u)$, then the moments of $\varphi_n(u)$ must approach the corresponding moments of $\varphi(u)$.[5] Letting $\mu_4(n)$ be the 4th moment of $\varphi_n(u)$ and μ_4 the corresponding moment of the normal curve, a necessary condition is that $\mu_4(n) \to \mu_4$ as $n \to \infty$, and $\mu_4 = 3$. The 4th moment of $\varphi_n(u)$ may be expressed simply in terms of the moment of $f_n(x)$. If the symbol E stands for expected value, the second and fourth moments of $f_n(x)$ are $E(x_n - \bar{x}_n)^2$ and $E(x_n - \bar{x}_n)^4$ respectively, and the relationship is then

$$\mu_4(n) = \frac{E(x_n - \bar{x}_n)^4}{[E(x_n - \bar{x}_n)^2]^2} = \frac{E\left[\sum_{i=1}^{n}(s_n - \bar{s}_n)\right]^4}{\left\{E\left[\sum_{i=1}^{n}(s_n - \bar{s}_n)\right]^2\right\}^2}. \tag{15}$$

Expanding the sums by the multinomial theorem and taking the expected value of each term we obtain

$$E(x_n - \bar{x}_n)^2 = \sum_{i=1}^{n} E(s_i - \bar{s}_i)^2 + 2\sum_{i<j=1}^{n} E(s_i - \bar{s}_i)E(s_j - \bar{s}_j) = \sum_{i=1}^{n} M_2(i), \tag{16}$$

where $M_2(i)$ is the second moment of $g_i(x)$ In a similar manner we have

$$\begin{aligned} E(x_n - \bar{x}_n)^4 &= \sum_{i=1}^{n} M_4(i) + 6\sum_{i<j=1}^{n} M_2(i)M_2(j) \\ &= \sum_{i=1}^{n} M_4(i) + 3\left[\sum_{i=1}^{n} M_2(i)\right]^2 - 3\sum_{i=1}^{n} M_2^2(i); \end{aligned} \tag{17}$$

[5] Uspensky, loc. cit., pages 383–388.

whence

$$\mu_4(n) = 3 + \frac{\sum_{i=1}^{n} M_4(i) - 3\sum_{i=1}^{n} M_2^2(i)}{\left[\sum_{i=1}^{n} M_2(i)\right]^2}. \tag{18}$$

Since a necessary condition for normality is that Lim $\mu_4(n) \to \mu_4 = 3$, the fraction in the above equation must in the limit approach zero. Substituting $M_2(i) = \frac{1}{3}\gamma_1$ and $M_4(i) = \frac{1}{9}\gamma_1^2 - \frac{1}{15}\gamma_2$, we find that this ratio reduces immediately in the limit to the condition $\Gamma = 0$.

5. **Application to the Distribution of Inversions.** A frequency table may be set up for the number of permutations of n objects that give rise to a fixed number of inversions. Three objects marked 1, 2, 3 may be permuted in 6 ways:

$$(123), (132), (213), (231), (312), (321).$$

If (123) is taken as standard position, the number of inversions associated with the above set to bring each one into standard position are respectively 0, 1, 1, 2, 2, 3. Thus we pass from (321) to (123) by the following three inversions or adjacent interchanges: (312), (132), (123). Among the six permutations there is one giving rise to 0 inversions, two having 1 inversion, two having 2 inversions, and one having 3 inversions

The distribution of inversions finds its application in a test of significance. The standard position is taken as a *hypothesis* of rank order, and the difference between an observed set of ranks and the hypothetical one is measured by the number of inversions. The distribution may then be used for finding the probability of obtaining by chance the number of inversions found, or less. For a moderate number of ranks (six or more), the distribution of inversions may be approximated by a normal curve. We shall show that as the number of ranks is increased, the normalized distribution of inversions approaches the normal distribution. The distribution of inversions of 1, 2, 3, 4, objects will be found in the table below.

Inversions: x	0	1	2	3	4	5	6
$1 \cdot f_1(x)$	1						
$1 \cdot 2 \cdot f_2(x)$	1	1					
$1 \cdot 2\ 3 \cdot f_3(x)$	1	2	2	1			
$1 \cdot 2 \cdot 3 \cdot 4\ f_4(x)$	1	3	5	6	5	3	1

By induction one may show that the following relationships hold between successive distributions:

$$
\begin{aligned}
f_2(x) &= \tfrac{1}{2}[f_1(x-0) + f_1(x-1)], \\
f_3(x) &= \tfrac{1}{3}[f_2(x-0) + f_2(x-1) + f_2(x-2)], \\
&\vdots \\
f_n(x) &= \frac{1}{n}[f_{n-1}(x-0) + f_{n-1}(x-1) \\
&\qquad + f_{n-2}(x-2) + \cdots + f_{n-2}(x-n+1)].
\end{aligned} \tag{19}
$$

Since this satisfies the basic recurrence formula (2), where $a_n = n - 1$, we may find out whether the normalized distributions of inversions approaches $\varphi(u)$. With $\gamma_n = n^2 - 1$ the condition $\Gamma = 0$ becomes $\operatorname{Lim}_{n\to\infty} \sum_{i=2}^{n} (i^2-1)^2 \Big/ \Big[\sum_{i=2}^{n} (i^2-1)\Big]^2$. The numerator sums to a polynomial of the 5th degree in n, while the brackets of the denominator sums to a 3d degree polynomial, which after squaring is of the 6th degree; so that as $n \to \infty$ we have in the limit $\Gamma = 0$. Thus the normalized distribution function of the inversions of n objects approaches $\varphi(u)$ as $n \to \infty$.

Equations (12) and (16) permit us to find the mean and standard deviation of the distribution of the inversions of n objects:

$$
\begin{aligned}
\bar{x}_n &= \tfrac{1}{4}n(n-1), \\
\sigma_n^2 &= \frac{1}{72}n(n-1)(2n+5).
\end{aligned} \tag{20}
$$

The sequence of binomial coefficients, and the distribution of inversions are examples of sequences that satisfy recurrence relation (2); it should be noted that their respective values of γ_n, ($\gamma_n = 3/4$ or $\gamma_n = n^2 - 1$), may be considered as *bounded* between two polynomials of the same degree in n. Whenever this is true the condition $\Gamma = 0$ will hold and $\varphi_n(u)$ will approach $\varphi(u)$. On the other hand, if for example, $\gamma_n = r^n$, then $\Gamma \neq 0$ and $\varphi_n(u)$ does not approach $\varphi(u)$.

6. **Smoothing Formulas.** The general recurrence formula (6),

$$
f_n(x) = \sum_{i=-\infty}^{+\infty} g_n(i) f_{n-1}(x-i),
$$

may be considered as a linear smoothing formula. For example, we may obtain the usual three point smoothing formula based on binomial coefficients for smoothing a distribution $f_1(x)$ into $f_2(x)$ by setting in the above equation $n = 2$, $g_2(i) = \tfrac{1}{4}\binom{2}{i+1}$ for $-1 \leq i \leq +1$, and $g_2(i) = 0$ for $i < -1$ or $i > +1$. Thus

$$
f_2(x) = \tfrac{1}{4}[f_1(x+1) + 2f_1(x) + f_1(x-1)]. \tag{21}
$$

From considerations found in Section 2, we see that if a variable x_1 has for distribution $f_1(x)$ and a variable s_2 has for distribution $g_2(x)$, then their *sum* $s_2 + x_1$ has for distribution function the smoothed distribution $f_2(x)$. From this point of view, the smoothed distribution $f_2(x)$, obtained by applying a linear smoothing formula, is a "cross" between the original unsmoothed distribution $f_1(x)$ and the artificial weight distribution $g_2(x)$.

Often a smoothing formula is used several times; first on the original distribution, then on the smoothed distribution, and then sometimes on the smoothed smoothed distribution. *If a linear smoothing formula is thus iterated* 1, 2, $3, \cdots, n, \cdots$ *times, the sequence of smoothed distributions obtained upon normalization approaches* $\varphi(u)$. This may easily be demonstrated by showing that Liapounoff's condition for normality, $\Gamma' = 0$, is satisfied. Since in this case the weight distribution $g_n(i)$ is the same for all $n \geq 2$, the corresponding moments of these distributions must all be equal; thus we may write $M_4(n) = M_4(2)$ and $M_2(n) = M_2(2)$ where $n \geq 2$. Substituting in (10), we obtain for Γ'

$$\Gamma' = \operatorname*{Lim}_{n\to\infty} \frac{M_4(1) + (n-1)M_4(2)}{[M_2(1) + (n-1)M_2(2)]^2}, \tag{22}$$

where $M_2(1)$ and $M_4(1)$ are the 2d and 4th moments of the unsmoothed distribution $f_1(x)$. The mean value $\bar{x}_n$ and the standard deviation σ_n of the distribution $f_n(x)$ formed by iterating a smoothing formula $n - 1$ times are easily shown to be

$$\begin{aligned} \bar{x}_n &= \bar{x}_1 + (n-1)\bar{s}_w, \\ \sigma_n^2 &= \sigma_1^2 + (n-1)\varphi_w^2, \end{aligned} \tag{23}$$

where $\bar{x}_1$ and σ_1 are the mean and standard deviations of the original unsmoothed distribution, and where $\bar{s}_w$ and σ_w are the mean and standard deviation of the weight distribution $g_2(i)$.

The linear smoothing formula is used in practical work to smooth data Successive application of one or many such linear formulas will usually smooth *any* set of values to the normal curve of error The above section serves as a warning of what is introduced by the use of such methods

It is a pleasure to acknowledge the helpful criticisms and advice of Dr W E. Deming in the preparation of the manuscript.

WASHINGTON, D C.

THE LENGTH OF THE CYCLES WHICH RESULT FROM THE GRADUATION OF CHANCE ELEMENTS

By Edward L. Dodd

1. **Introduction.** Eugen Slutzky[1] found that under certain conditions repeated summations of chance elements lead to a sinusoidal configuration. Generalizations were made by V. Romanovsky.[2] A more recent paper by Slutzky[3] has appeared, summarizing his original Russian memoir, and making extensions. Contributions to this subject have also been made by H. E. Jones,[4] E. J. Moulton,[5] and A. Wald.[6]

Readers who wish to get into touch with recent literature on periodicity are referred to two excellent books, that of Karl Stumpff[7] with bibliography of 319 references, and that of Herman Wold,[8] with bibliography of nearly 70 references.

In this paper, I deal with the wavy configuration resulting from a *single* application of a specified graduation formula. For this purpose, only linear operators are considered. For actual graduation it is customary to require that the sum of the coefficients or "weights" be equal to unity. But for the present purpose, this requirement is irrelevant. For example, summing and averaging are here essentially identical. The graduation formula considered may or may not be the combination of simple summations or averages. Indeed, formulas preferred by actuaries and statisticians include terms with *negative* coefficients; and thus involve an operation other than addition. F. R. Mac-

---

[1] Eugen Slutzky, "Sur un théorème limite relatif aux series des quantités éventuelles." *Comptes Rendus*, Vol 185 (1927) pp 169-171.

[2] V. Romanovsky, "Généralisations d'un théorème de M E. Slutzky." *Comptes Rendus*, Vol 192(1931) pp. 718-721. "Sur la loi sinusoidale limite." *Rendiconto Circolo Mathematico di Palermo*, Vol. 56 (1932) pp. 82-111 "Sur une généralisation de la loi sinusoidale limite." Ibid, Vol 57 (1933) pp. 130-136.

[3] E. Slutzky, "The summation of random causes as a source of cyclic processes" *Econometrica*, Vol. 5 (1937) pp. 105-146.

[4] H. E Jones, "The theory of runs applied to time series," *Report of Third Annual Conference of Cowles Commission for Research in Economics* (1937) pp. 33-36. This abstract itself does not include reference to repetitions, mentioned by Moulton and Wald

[5] E J. Moulton, "The periodic function obtained by repeated accumulation of a statistical series *American Mathematical Monthly*, Vol. 45 (1938), pp. 583-586.

[6] A Wald, "Long cycles as a result of repeated integration." *American Mathematical Monthly*, Vol. 46 (1939), pp. 136-141.

[7] Karl Stumpff, *Grundlagen und Methoden der Periodenforschung*, Berlin, 1937, Julius Springer.

[8] Herman Wold, *A Study in the Analysis of Stationary Time Series* Uppsala, 1938, Almqvist and Wiksells.

aulay[9] gives a chart of 24 weight diagrams. Of these only the first four are without negative coefficients.

Of course, the "waves" produced are irregular, and the difficulty of defining a cycle-length confronts us. The apparently naïve definition of a cycle-length as the average distance between suecessive maxima (or minima) is, I believe, worth consideration as a rough first approximation of the cycle length for graduated values delivered by formulas with negative coeffieients or by those involving at least triple summations. But the cyele length thus determined is somewhat too short; for, slight undulations will occur—Slutzky calls them "ripples"—which should be eliminated if we want a cycle-length *intuitionally reasonable*. On the other hand, the eycle-length defined as the average distance between alternate intersections of the graduated curve with the base line is likely to be decidedly too long,—as illustrated by Slutzky's Figure 2 (loc. cit., p. 109) which exhibits 1,000 graduated items, with 41 marked maxima and 41 marked minima—after *elimination* of what he considers ripples—but with only 23 up-crossings and 23 down-crossings of the base line. I indicate in what follows an analytic method for removing ripples. And I describe *several methods* for obtaining a number which might be called a cycle-length. Often these seem to *cluster about a central value*, which appears to me to be a reasonable estimate of the "*length of the cycle*" created by the *specified graduation formula*.

The *theory* to be presented here assumes that the chance elements are *normally distributed* about zero with constant variance. But the data used by Slutzky came from lottery drawings, with a "rectangular" distribution; and for ehecking I have likewise used *rectangular distributions*; mainly, three sets of 600 numbers each, taken from the tenth figures of logarithms in the Vega Tables. It is known, however, that the average of a few elements distributed rectangularly is nearly normal. From many tests that I have made, it would seem that rectangular distributions react as if normal. To illustrate: When normal data are given a twelve-fold summation or averaging by twos, the probabilities that at a specified point there will be an uperossing of the base line, a maximum, or an inflection from concave to convex are respectively, 0.0628, 0.106, and 0.134. These numbers multiplied by 100 give 6.28, 10.6, and 13.4, as the expected number of occurrenees per hundred graduated values. Slutzky exhibits in Figure 4 (loc. cit., p. 111) 100 ordinates as the result of applying to lottery drawings 12-fold summation by twos. The figure shows 6 or 7 up-crossings, ten maxima, and 13 or 14 sueh infleetions—in close agreement with *expectations based upon normal distributions*.

2. **Derivation of Probabilities and Comparison of Actual with Expected Occurrences.** A "cycle length" is first conceived of as the *reciprocal of a relative frequency or probability*. Thus, if the probability that a graduated value will

[9] F R. Macaulay, *The Smoothing of Time Series* Publications of the National Bureau of Economic Research, incorporated, No. 19 (1931). See pp 77–79.

be a maximum is 0.05, we expect 5 maxima per hundred graduated values, making the "cycle length" for maxima equal to 20. It will be recalled that if p is the probability of an occurrence of an event in a single trial, then in s trials the expected number of occurrences is sp, whether the trials are *independent or not.*

It is assumed that the *data,* $x_1, x_2, \cdots$ are *independent and normally distributed about zero with constant variance* V. Then any linear function

$$(1) \qquad y_r = a_{-m}x_{r-m} + \cdots + a_0x_r + a_1x_{r+1} + \cdots + a_mx_{r+m}$$

is likewise normally distributed about zero; and the variance of y_r is $V = \Sigma a_i^2$.

(a) *Probabilities When Two Conditions Are Imposed.* Consider first the "planes" $y_{r-1} = 0$ and $y_r = 0$, each in $2m + 1$ dimensions; and jointly in $2m + 2$ dimensions. They form four "dihedral" angles. Let

$$(2) \qquad \theta = \text{angle between } y_{r-1} = 0 \quad \text{and} \quad y_r = 0,$$

the inside points $(x_{r-m-1}, \cdots, x_{r+m})$ being such that $y_{r-1} < 0$, and $y_r > 0$. Now, an orthogonal transformation or "rotation "leaves invariant this angle θ and also the normal probability function:

$$(3) \qquad \text{Probability} = \text{Const.} \cdot \exp[-\Sigma x_i^2/2V].$$

The angle θ may be found[10] from

$$(4) \qquad \cos\theta = \sum_{i=-m}^{m-1} a_ia_{i+1} \Big/ \sum_{-m}^{m} a_i^2.$$

Let us think of the rotation which carries the intersection of the planes into the "vertical" position. To find the probability that $y_{r-1} < 0$ and $y_r > 0$, we integrate over all $2m + 2$ dimensional space which lies between the two planes in the dihedral angle thus characterized. For $2m$ of such variables, the integration extends from $-\infty$ to $+\infty$ yielding unity as a factor. If u and v are the two variables that remain, then we are to find the volume of that portion of the solid

$$(5) \qquad z = (1/2\pi V) \exp[-(u^2 + v^2)/2V]$$

which lies between two vertical planes through the origin making the angle θ with each other. Then,

$$(6) \qquad \text{Probability of up-crossing} = \theta/360°.$$

$$(7) \qquad \text{Cycle length for up-crossing} = 360°/\theta.$$

Let

$$\Delta y_r = y_{r+1} - y_r.$$

---

[10] D. M. Y. Sommerville *An introduction to the Geometry of N Dimensions* Methuen and Co., Ltd., London, 1929. See p. 76.

Then y_r is a maximum if $\Delta y_{r-1} > 0$ and $\Delta y_r < 0$. Suppose

(8) $\theta_1 =$ angle between $\Delta y_{r-1} = 0$ and $\Delta y_r = 0$,

inside points making $\Delta y_{r-1} > 0$ and $\Delta y_r < 0$. Then

(9) Probability for maximum at $y_r = \theta_1/360°$

(10) Cycle length for maxima $= 360°/\theta$.

The same equations apply to minima; since for minima we simply reverse the two foregoing inequalities, and pass to the equal "vertical" dihedral angle. Likewise, from $\Delta^2 y_{r-1} < 0$ and $\Delta^2 y_r > 0$ we obtain an angle θ_2 such that $\theta_2/360°$ is the probability for change of inflection from concave downward to convex downward. This is also equal to the probability for change of inflection from convex to concave. Such changes of inflection have some interest on their own account and in checking; but do not seem to have any direct bearing upon the main problem under discussion here.

(b) *Probabilities When Three Conditions Are Imposed.* We consider now the *elimination of ripples.* To make y_r a maximum, two linear conditions are required. A third linear condition such as $y_r > \frac{1}{2}(y_{r-k} + y_{r+k})$, or simply $y_r > y_{r+k}$, with $k > 1$, will remove some ripples. Suppose we have given three planes through the origin,

$$\begin{aligned} a_1x_1 + a_2x_2 + \cdots + a_nx_n &= 0, \\ (11) \qquad b_1x_1 + b_2x_2 + \cdots + b_nx_n &= 0, \\ c_1x_1 + c_2x_2 + \cdots + c_nx_n &= 0. \end{aligned}$$

The angles between these planes in pairs are

$$(12) \quad \cos\alpha = \frac{\Sigma b_i c_i}{(\Sigma b_i^2 \cdot \Sigma c_i^2)^{\frac{1}{2}}}; \qquad \cos\beta = \frac{\Sigma a_i c_i}{(\Sigma a_i^2 \cdot \Sigma c_i^2)^{\frac{1}{2}}}; \qquad \cos\gamma = \frac{\Sigma a_i b_i}{(\Sigma a_i^2 \cdot \Sigma b_i^2)^{\frac{1}{2}}}$$

In general, eight-trihedral angles are thus formed at the origin; since we may take acute angles for α, β, and γ or their supplements. By an orthogonal transformation or "rotation about the origin" we are led to the three dimensional problem of finding the portion of a sphere lying in a specified spherical pyramid with base a spherical triangle, ABC, having spherical excess $E = A + B + C - 180°$. Now the spherical surface is 4 great circles or 720°. Hence, for a maximum, subject to an additional linear homogeneous inequality,

(13) Probability of conditioned maximum $= E/720°$

care having been taken to enter the proper trihedral angle.

(c) *Probabilities When Four Conditions Are Imposed.* To avoid complexities involved in the use of four intersecting planes, the following expedient was employed. Consider a set of values of y_r such that this y_r is a maximum. Among these there is theoretically a certain fraction or proportion p at which also

$y_r > y_{r+k}$, with $k > 1$, and the same proportion p satisfying $y_r > y_{r-k}$. Let p' be the proportion satisfying both inequalities Then $1 - p' \leqq 1 - p + 1 - p$ leads to

$$p' \geqq 2p - 1 = p^2 - (1 - p)^2. \tag{14}$$

If p is fairly close to unity; a good approximation for p' would seem to be

$$p' = p^2. \tag{15}$$

This p^2 would have been exact for p', had the graduated values been independent. That p' is here only slightly above $2p - 1$ seems likely, from the graduations that I have examined; for, the failure of one of the inequalities $y_r > y_{r+k}$ or $y_r > y_{r-k}$ was seldom accompanied by the failure of the other.

For graduations with the Spencer 21-term formula, when $k = 5$, $p = 0.936$, and $(1 - p)^2 = 0.0041$, which is fairly small. In practice, we would find in this case directly $P = 0.07125$ = probability of a maximum; $Pp = 0.0668$ = probability of a maximum at y_r with $y_r > y_{r+5}$ Then the probability Pp' of a maximum at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$ would have as lower bound $2Pp - P = 2(0\ 0668) - 0.07125 = 0.06235$.

But a closer approximation to the actual value would seem to be $Pp^2 = (Pp)^2/P = (0.0668)^2/0.07125 = 0.0626$

This would give a cycle length of $1/0\ 0626 = 15\ 97$.

(d) *Indications from Correlation Theory.* We may also attempt to estimate a cycle length with the aid of correlation theory If for graduation, we use a linear operator with coefficients proportional to successive ordinates of a cosine curve with a specified period, it is, I presume, fairly well known that the graduated values tend to exhibit the period of that cosine curve Moreover, this quasi period may be induced very strongly with the use of formulas which represent "damped vibration" as shown by H. Labrouste[11] and Mrs Labrouste Now many standard graduation formulas have plots resembling somewhat damped vibration Here, the central symmetrical arch leading down to the lowest negative terms on each side is usually large in comparison with the flanking waves. Now for a *strict cosine* curve of period $2k$, the coefficient of correlation of y_r and y_{r+k} is -1, at least theoretically. For *chance* material y_r, with mean zero and constant variance, the coefficient of correlation between y_r and y_{r+j} is defined in terms of expected values, thus:

$$\rho_j = E(y_r y_{r+j})/E(y_r^2). \tag{16}$$

For graduated values, y_r, we might then *seek the value j which will make ρ_j as close to -1 as possible.* But for most common graduation formulas, ρ_j does not approximate -1. This difficulty, however, disappears if the graduation

[11] H. and Mrs Labrouste, "Harmonic analysis by means of linear combinations of ordinates," *Terrestrial Magnetism and Atmospheric Electricity*, Vol. 41 (1936) pp 15–28. See pp 17, 18.

formula is properly centered. In a Fourier series, there is a constant term, which is of no significance in indicating oscillations, and is sometimes eliminated. The analogous modification for a linear graduation formula with n coefficients—of which the sum is unity—would seem to be the subtraction of $1/n$ from each coefficient, forming what I regard as a *residual. For this residual, negative correlations of substantial size appear.* And that j with which the numerically largest negative correlation ρ_j is associated may be considered as indicating a half-cycle length.

In the case of the Spencer 21-term formula, $j = 8$, making cycle-length $= 16$, just about identical with the cycle length for maxima at y_r with $y_r > y_{r-5}$ and $y_r > y_{r+5}$.

(e) *The period of a Closely Fitting Cosine Curve.* By another route, also, we may approach the problem of associating with a specified linear graduation a number as the length of induced cycles. We shall consider here only those formulas in which the coefficients are symmetrical with respect to the center. In equation (1), this means that $a_{-j} = a_j; j = 1, 2, \cdots, m$. Suppose now that the x's are no longer chance elements, but are the successive terms of a cosine curve with period k. That is:

$$x_r = \cos(r\theta + \alpha); \qquad \theta = 2\pi/k = 360°/k. \tag{17}$$

Then, if $a_{-j} = a_j$, it follows that

$$a_{-j}x_{r-j} + a_j x_{r+j} = 2a_j \cos j\theta \cdot \cos(r\theta + \alpha) \tag{18}$$

Then, from (1),

$$y_r = C \cos(r\theta + \alpha), \tag{19}$$

where C is independent of r. For a given graduation formula, with a's specified, this C depends upon θ, or we may say, upon $k = 360°/\theta$. *We may regard the graduation formula y_0 as "fitting best" the curve* cos $[r(360°)/k]$ *when k is so chosen as to give to C a largest value.* The presumption is that the graduation formula will curl chance data up into cycles in about the same fashion as a cosine curve to which it is closely akin. The actual period of this closely fitting cosine curve may then be taken as the quasi-period or "cycle-length" of the graduation formula.

If, relying upon intuition, we were to select a cosine curve to fit a given graduation formula, we might easily decide to disregard the small waves that usually flank the central arch, and to take a cosine curve with a span—distance between minima—equal to the span of this central arch. *In fact, this span gives, I believe, a good first estimate of the cycle length of the induced waves. This first estimate seems, however, a trifle too small.*

3. Size of Ripples, Simple Summation, Variability, and Height of Waves.

(a) *Size of Ripples.* In the use of $y_r > y_{r+k}$ to remove ripples, what integer should we take for k? The dividing line between ripples and waves is of course

arbitrary. As Figure 2, p. 109, Slutzky exhibits 1,000 graduated values from two-fold summations by 10, with ripples removed. He states (p. 119): "maxima and minima with amplitudes of ten units or less being discarded as ripples." For this double summation, I find that the probability that y_r will be a maximum with $y_r > y_{r+10}$ and $y_r > y_{r-10}$ is approximately 0.0437. Among 1,000 graduated values, 43.7 such maxima would then be expected. Slutzky marks with arrows the 41 maxima which remain after the elimination of what he regards as ripples. The reciprocal of 0.0437 gives 22.9 as cycle length. Then $k = 10$ is less than half this cycle-length. For standard graduation formulas, it would seem likely that a *value of k about one-third the span of its central arch would eliminate fairly well the inconsequential fluctuations*; and likewise for graduations, with coefficients forming an arch with nearly horizontal ends, like twelve-fold summation by twos, with arch span 12. For this twelve-fold summation, I find that 0.0831 is the probability that a maximum will occur at y_r, with $y_r > y_{r+4}$ and $y_r > y_{r-4}$, giving 8.31 such maxima per hundred graduated values. Slutzky's Figure IVa shows eight such maxima, and two ripples.

(b) *Simple Summation.* I shall not discuss in detail the cycles produced by simple summation or averaging. Formulas for probability here are relatively simple. Thus, for the sum or average of n normal chance data, the probability of a maximum is 1/4, irrespective of the value of n. This appears to be about valid for rectangular data if we count the weak maxima *A simple average of chance data, however, seems to inherit largely the chaotic character of the present data. But some sinuosity is, after all, implanted.*

(c) *Variability.* A general discussion of the variability of induced waves is beyond the scope of this paper. However, I record a numerical result. For the Spencer 21-term graduation formula, the probability of a maximum is 0.07125. Among 580 graduated values, then, 41.3 maxima would be expected. Actually, 42 maxima were found. Now, if $n - 1$ points are placed "at random" on a line of unit length—here dx is the probability that a point will fall in an interval of length dx—then the expected value[12] of the sum of the squares of the resulting n segments is $2/(n + 1)$. Thus, if 42 points are placed at random on an interval of 580 units, the expected sum of the squares of the segments is $(2/44)\ (580)^2 = 15{,}290.9$. But, if the points are placed at equal intervals, this sum of squares takes its least value, $(580)^2/43 = 7{,}823.3$. Then, $15{,}290.9 - 7{,}823.3 = 7{,}467.6$. On the other hand, the 42 maxima among *Spencer graduated values* gave segments for which the sum of the squares was 8,656.5; that is, *only* 833.2 *in excess of the above* 7,823.3, *which represents perfect periodicity for maxima.* Of course, this excess of 833.2 indicates considerable departure from perfect periodicity; but it is nowhere near the 7,467 6 to be expected from a random distribution of points. In spite of irregularities, *the sinusoidal character of graduated values is conspicuous.*

(d) *The Height or Amplitude of Induced Waves.* While our chief interest

[12] W. Burnside, *Theory of Probability*, Cambridge University Press, 1928. See p. 71.

here lies in what is called the *length* of a cycle, a brief reference may well be made to the amplitude or *height* of the induced waves. The operation of the linear function y_r in (1) upon data with variance V yields graduated values with variance $V\Sigma a_i^2$. This particular statement does not require the assumption of normality Thus the Spencer 21-term formula is expected to produce graduated values with a standard deviation 37.8% of that of the data. This represents some reduction, of course; but, nevertheless, the *"waves" stand out in bold relief. They are not diminutive.*

4 **Data and Graduations Examined.** Slutzky's graduations, exhibited in *Econometrica*, Vol. 5, have already been mentioned. Three sets of chance data were graduated by students at the University of Texas, Mr. Victor W. Pfeiffer in 1936, Mr C. M Tolar and Miss Anna Velma Martin, in 1938, to make tests with regard to smoothing coefficients,[13] the results appearing in M.A. theses. The data were figures in the tenth place of the Vega logarithm tables, 600 numbers in each set, as follows· Logarithms from 200 to 799; logarithms of cosines of angles from 0° to 59° 54′, by intervals of 6′; logarithms of sines of angles from 6′ to 60° by intervals of 6′. The graduation formulas used were all symmetric, with $a_{-i} = a_i$. Mr Pfeiffer used the Spencer 21-term formula, with coefficients 1/350 of.

$$-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, 60, 57, \text{etc.}$$

The other two formulas used were 11-term formulas which I devised, correct to third differences, and with fourth differences rather small, described by: $-1.13\,D^4$ and $-0\,97\,D^4$, where $D = \log_e E$ (see Henderson, loc. cit., pp. 26–37); as compared with $-5.4\,D^4$ for Woolhouse 15-term, and $-12.6\,D^4$ for Spencer 21-term. These two 11-term formulas are:
(i) Averaging by twos, threes, and fours, applied to (1/12) (−4, 3, 14, 3, −4) yielding (1/288) (−4, −9, 3, 36, 73, 90, 73, 36, 3, −9, −4);
(ii) Triple averaging by threes, applied to (1/10) (−3, 2, 12, 2, −3) yielding (1/270) (−3, −7, 0, 29, 71, 90, 71, 29, 0, −7, −3). From part of the foregoing data, also, I made other graduations to check certain probabilities.

5. **Cycle Lengths for the Spencer 21-Term Graduation Formula.** All the various determinations of cycle length mentioned in the foregoing were applied to the Spencer formula, and to some other formulas The results obtained for the Spencer formula seem representative, and will be given here in detail. Our main conception of a cycle-length is that it is the reciprocal of a probability or relative frequency. The probability of a minimum is the same as that of a maximum; of a down-crossing of the base line, the same as that of an up-crossing Probabilities are listed that the representative ordinate y_r will be a maximum—

[13] Robert Henderson, *Graduation of Mortality and Other Tables*, Actuarial Society of America, New York, 1919, p. 34.

with or without further restrictions. The probability is given for an up-cross at the representative abscissa x_r. In the table which follows, a middle entry for a cycle length of 16 is obtained from the "residual" described in (d) of Section 2.

The Expected Length of Cycles Produced When Normal Chance Data Are Graduated by the Spencer 21-term Formula in Accordance with Various Specifications for the Cycle

Specification	Probability	Cycle-Length
Maximum at y_r	0.07125	14.0
Maximum at y_r with $y_r > y_{r+5}$	0 0668	15.0
Maximum at y_r with $y_r > \frac{1}{2}(y_{r-7} + y_{r+7})$	0.0657	15.2
Maximum at y_r with $y_r > y_{r+5}$, and $y_r > y_{r-5}$	0.0626	16.0
By use of "residual". (See 2(d))		16.0
Maximum at y_r with $y_r > y_{r+7}$	0.0623	16.1
Period of "best fitting" cosine curve. (See 2(e))		16.7
Maximum at y_r, $y_r > 0$. (Or: $y_r >$ Mean y_r)	0 0591	16.9
Maximum at y_r with $y_r > y_{r+7}$ and $y_r > y_{r-7}$	0.0545	18 3
Up cross at x_r	0.0469	21.3

The foregoing exhibit seems to suggest a cycle length of something like 16 *for the cycles created by the operation of the Spencer* 21*-term formula upon chance data.* This is just about the reciprocal of the probability that a maximum will occur at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$. If 16 is thus set up as the standard wave length, a wave of 10 units extending from x_{r-5} to x_{r+5} would not be regarded as insignificant.

Now 16 is also the interval between the outermost low coefficients, -5, in the Spencer formula. The plot of a curve through ordinates equal to the Spencer coefficients would probably make the central arch have a span of about 15. This 15 seems a little too small as a representative of cycle lengths obtained by the foregoing different methods.

From the theory set forth, 0.0626 is the probability that a maximum will occur at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$. Then among 580 graduated values, 36 3 such maxima would be expected Among the Pfeiffer graduated values 38 were actually found.

6. **Comparative Results of Seven Graduation Formulas.** An exhibit will now be made of results obtained from seven graduation formulas Of these, the simplest is double averaging or summation by tens, with coefficients forming a

triangular arch, with a "span" which will be set down here as 18 Next in order of simplicity—avoiding negative coefficients—is 12-fold averaging by twos. Probabilities are given that a maximum will occur at a point y_r, with $y_r > y_{r-k}$ and $y_r > y_{r+k}$ for what seems to be appropriate values of k. In the five cases where graduations were made, the number of the maxima of specified character actually found are set down in line with their expected values. Also the span of the central arch is compared with cycle lengths.

Macaulay (loc. cit., pp. 73, 74) mentions favorably a 43-term formula obtained as follows: Summation by 8, by 12, doubly by 5, applied to weights: +7, −10, 0, 0, 0, 0, 0, 0, +10, 0, 0, 0, 0, 0, 0, −10, +7. This is the longest formula to be considered here.

As noted before, my theory is based upon the assumption of a normal distribution for the data. The data actually tested had a rectangular distribution. Nevertheless, close agreement was found between the expected number of maxima and the number actually found.

Results of Applying Seven Graduation Formulas to Chance Data. Comparison of the Expected Number of Conditioned Maxima with the Actual Number Found Among Graduated Values, and Comparison of Cycle Length with Span of Central Arch

(1) Graduation Formula	(2) k	(3) Probability Max at y_r $y_r > y_{r-k}$ $y_r > y_{r+k}$	(4) Number of Graduated Items, y_r	(5) Expected Number of Such Maxima	(6) Actual Number of Such Maxima	(7) Cycle as 1/(3)	(8) Span of Central Arch
11-term by Tolar	3	0 110	590	64 9	67	9 09	8
11-term by Martin . .	3	0.114	590	67 3	65	8 77	8
13-term $(2)^{12}$ by Slutzky . . .	4	0.0831	100	8.31	8	12.0	12
19-term $(10)^2$ by Slutzky	10	0.0437	1,000	43.7	41	22 9	18
21-term Spencer by Pfeiffer.. . .	5	0.0668	580	36.3	38	16 0	15
29-term Kenchington .	8	0 0428				23.4	20
43-term Macaulay . . .	9	0.0389				25 7	22

7 **Summary.** E Slutzky found that the summing of chance data resulted in series of numbers with something like a cyclic appearance,—this being intensified by repetition of the summing. Slutzky and others have proven limit theorems. In this paper, I study the effects of a single application of a graduation process upon chance data. The most acceptable graduation formulas contain negative coefficients, and thus involve something more than repeated summations. Several methods are discussed for assigning to a given graduation formula a number as the length of the cycles it tends to produce. One of the most satisfactory of these is in line with the suggestion of Slutzky that before counting maxima, any insignificant "ripples" should be eliminated. The proba-

bility is found that a graduated value y_r should be a maximum—greater than the two adjacent values y_{r-1} and y_{r+1}—with the further condition that for some appropriate k, y_r shall be greater than y_{r-k} or y_{r+k} or both. The reciprocal of this latter probability is suggested as the length of the cycle which the given graduation tends to implant in the graduated values.

THE UNIVERSITY OF TEXAS.

ON THE DISTRIBUTION OF THE "STUDENT" RATIO FOR SMALL SAMPLES FROM CERTAIN NON-NORMAL POPULATIONS[1]

By H. L. Rietz

Much of interest in the theory and practice of statistical methods has been developed around the distribution function,

$$(1) \qquad \frac{\Gamma(N/2)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{N-1}{2}\right)(1+z^2)^{N/2}}$$

of the "Student" ratio, $z = \dfrac{\bar{x} - m}{s}$, where $\bar{x}$ denotes the mean, s the standard deviation of a sample of N items, say $x_1, x_2, \cdots, x_N$, taken at random from a normally distributed parent population of mean, m.

The investigations of certain non-normal parent distributions by Shewhart and Winters [1], Rider [2], E. S. Pearson [3], M S. Bartlett [4], and R. C. Geary [5] indicate that applications of the "Student" theory give more satisfactory results than the classical theory for a considerable variety of non-normal parent distributions, but some of these investigators find that the theory fails in certain cases to describe the facts to an extent that suggests further experimental sampling investigations along this line whenever suitable data are available. Others infer that a completely satisfactory analysis of the position of the "Student" z-test will be possible only if the theoretical distribution of z in samples from the non-normal distribution in question becomes known. Several of the above named statisticians have attributed the failures of the distribution (1) to describe their data, in large part, to the correlation between $x = \bar{x} - m$ and s For this reason, there is considerable interest in the degree of correlation between $x = \bar{x} - m$ and s, and especially in the nature of the regression of s or of s^2 on x.

The present paper gives an analysis of data obtained by experimental sampling from two non-normal distributions whose sources we shall now describe. The parent distributions with which the paper is concerned are theoretical distributions resulting from certain urn schemata devised [6] by the writer some years ago

In 1925, Leone E Chesire, in an unpublished thesis for the degree, Master of Science, at the University of Iowa, obtained data by experimental sampling, that seem to be appropriate material for a study of the correlation of mean and standard deviation for small samples from certain non-normal distributions

One of the original bivariate parent populations, whose marginal totals we are

[1] Presented in part before the American Mathematical Society under a somewhat different title, November 26, 1937.

using, exhibited linear regression while the other exhibited non-linear regression. For convenience in distinguishing between the two cases, we shall speak of material from the linear case as Case I and that from the non-linear case as Case II. After devising a scheme for drawing pairs of variates at random, 5,000 pairs were drawn in sets of five for each of the two cases.

While the primary purpose of this experimental sampling was to study the distributions of means, standard deviations, and correlation coefficients [7] for small samples from the non-normal populations, we have as a by-product, in the marginal totals of the correlation tables, four sets of 1,000 pairs of means and standard deviations. However, since three of the four sets of marginal totals of the two theoretical parent correlations tables are alike, we have actually only two significantly different sets to consider.

Case I. For the case of linear regression of y on x in the bivariate parent population, the parent distribution from the marginal totals may be very simply described by showing the frequency distribution in Table 1.

TABLE 1

Sums in second throw of dice-values of stochastic variable	2	3	4	5	6	7	8	9	10	11	12
Frequency	6	12	18	24	30	36	30	24	18	12	6

The moment coefficients and β's which characterize the distribution given in Table 1 are:

$$\text{Mean} = 7, \quad \mu_2 = 5\tfrac{5}{6}, \quad \mu_3 = 0, \quad \mu_4 = 80.5, \quad \beta_1 = 0, \quad \beta_2 = 2\tfrac{844}{1775}.$$

Each of the 1000 sets of five drawn from the distribution in Table 1, yields a mean $\bar{y}$ and a standard deviation, s_y, which we shall denote by w to make our notation simpler to write. Table 2 is the correlation table of the pairs $(\bar{y}, w)$. The correlation coefficient $r_{w\bar{y}}$, between mean $\bar{y}$ and standard deviation $s_y = w$ has a value

$$r_{w\bar{y}} = -0.020 \pm 0.021$$

which differs insignificantly from zero.

The uncorrected value of the correlation ratio of w on $\bar{y}$ is

$$\eta_{w\bar{y}} = 0.182.$$

When we remember that the correlation ratio is not free to vary in the negative direction from 0, and apply the Pearson correction [8] for this situation together with the "Student" correction [9] for grouping, we obtain for the corrected, $\eta_{w\bar{y}}$, the value 0.133.

It becomes fairly obvious that significant correlation exists and that the regression is non-linear. Indeed, it has been shown recently by Geary [5, pp. 178-9] that normality in the parent distribution is both a necessary and

TABLE 2

Correlation of mean $\bar{y}$, and standard deviation $s_y = w$, of samples of five items for Case I. Mean of $\bar{y}$'s $= \bar{\bar{y}} = 7.141$. Correlation coefficient $r_{w\bar{y}} = -0\,020 \pm 0.021$, $s_y = w = 2\,079$. Correlation ratio of w on $\bar{y}$, $\eta_{w\bar{y}} = 0.182$ (uncorrected)

$\bar{y}$

w	37	41	45	49	53	57	61	65	69	73	77	81	85	89	93	97	101	f_w
4.1								1		2	1							4
3.9							1	1			1	1						4
3.7				1				3		1			2					7
3.5			1					1	1	2	1	1	1					8
3.3					2	4	2	1		7	2	2	5	1				26
3.1				1	3	2	2	8	3	9	4	6	3					41
2.9					5	4	3	10	8	10	3	1	4	2				50
2.7			1	1		3	12	9	18	18	6	9			4			81
2.5			2		4	8	14	15	6	22	14	14	6	5	1			113
2.3				1	2	9	14	10	11	10	10	12	7	3				89
2.1			3	4	2	12	12	16	13	15	8	9	7	8	3			112
1.9	1		1	4	7	6	5	16	7	18	15	7	4	3	4	3		101
1.7				7	7	9	15	15	23	16	14	17	6	4		2		135
1.5		1	1	3	5	4	3	6	8	6	7	10	6	5	1	1	1	68
1.3			2	2	1	3	4	11	8	7	11	3	3	2	2			59
1.1				1	1	4	5	5	10	6	6	6	6	1			1	52
0.9			2	1			5	4	2		2	4		1		1		22
0.7					1		1		7	2	3	2						16
0.5								3	1	1	1		2		2			10
0.3									1									1
0.1									1									1
$f_{\bar{y}}$	1	1	13	26	40	68	98	135	130	152	109	104	62	35	17	7	2	1000

sufficient condition for the independence of the mean and standard deviation in samples.

Since the number of correlated items, $N = 1000$, is fairly large, we examine into the significance of $\eta_{w\bar{y}} = 0.182$ under the assumption that $N\eta^2_{w\bar{y}}$ is approximately distributed [10] as χ^2 with $a - 1 = 16$ degrees of freedom. This criterion gives odds in favor of significant correlation on approximately a 100 to 1 level of probability.

Next, the means of arrays, $\bar{w}_p$, were plotted to scale on Table 2 to give a general notion of the nature of the regression of $w = s_y$ on $\bar{y}$. The location of these means of arrays of w's affords at least a suggestion of parabolic regression [11] with the curvature concave downward as is to be expected when $\beta_2 - \beta_1 - 3 < 0$, where the β's relate to the parent distribution.

The next step taken was to analyze the variance, as indicated partly in Table 3, where w_i $(i = 1, 2, \cdots, N)$ denotes the stochastic variates, a the number of arrays of w's, $\bar{w}$ the mean of the N values of w_i, $n_p (p = 1, 2, \cdots, a)$ the number of variates in an array marked p, $\bar{w}_p$ the mean of the array marked p, and where the class interval in Table 2, is taken as the unit.

TABLE 3

	Sum of squares	
For deviations of means of arrays of w's	$\sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 = 380$	$a - 1 = 16$
For deviations of variates from the means of their arrays	$\sum\sum (w_i - \bar{w}_p)^2 = 11{,}098$	$N - a = 983$
Total	$\sum_{i=1}^{N} (w_i - \bar{w})^2 = 11{,}478$	$N - 1 = 999$

In the exhibit given in Table 3, we use the usual algebraic identity

$$(2) \qquad \sum_{i=1}^{N} (w_i - \bar{w})^2 = \sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 + \sum\sum (w_i - \bar{w}_p)^2,$$

where the double sum is made up of a sum of N squares.

By dividing the members of (2) by N, we have

$$(3) \qquad \frac{1}{N}\sum_{i=1}^{N} (w_i - \bar{w})^2 = \frac{1}{N}\sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 + \frac{1}{N}\sum\sum (w_i - \bar{w}_p)^2$$

The writer has used the identity (3) for many years in lectures to beginners in statistics in proving the equivalence of two definitions of the correlation [12] ratio and is strongly of the opinion that the equality in form (3) appeals more readily to the intuitions of many readers, because of their acquaintance with statements in the language of averages, than does the equivalent equality (2) in the language of sums of squares.

In an extended and more compact form, the analysis is shown in the standard form in Table 4.

TABLE 4

Variance	Degrees of freedom	Sum of squares	Mean square	z-test
Between arrays	16	380	23.75	$\frac{1}{2}\log_e 23.75 = 1.584$
Within arrays	983	11,098	11.29	$\frac{1}{2}\log_e 11\ 29 = 1.212$
Total	999	11,478		Difference = 0 372

When the sum of squares equal to 380 associated with variance between arrays is further analyzed into a part which could be represented by linear regression,

and a part which represents deviations of the calculated means of arrays of w's from a straight regression line of w on $\bar{y}$, the deviations being measured parallel to the w-axis, we find that the part of the amount 380 represented by linear regression is given by

$$Nr^2_{w\bar{y}}s^2_w = 1000\ (.00040)(11.487) = 4.3.$$

Since both $r = .020 \pm 0.021$ and the small value, 4.3, as part of the sum of squares amounting to 380, may well be regarded as sampling fluctuations, we revert to the figures in Table 3 and apply the Fisher z-test. It turns out that the correlation is significant on practically the 100 to 1 level of probability which conforms well with the above inference based on the assumption that $N\eta^2_{w\bar{y}}$ is distributed as χ^2, with $a - 1$ degrees of freedom

Next, we computed 1000 values of the "Student" ratio $z = (\bar{y} - 7)/w$, for Case I. One of these 1000 values was of the indeterminate form $\frac{0}{0}$. A frequency distribution of the 999 determinate ratios is shown in column (3), Table 5.

By grouping together the class frequencies at the tails of the theoretical distributions until each of the end class frequencies is not less than 5, and calculating χ^2 for the observed distribution in column (3) in comparison with the theoretical distribution in column (6) as found from the "Student" theory in samples of 5 items from a normal distribution, we obtain $\chi^2 = 3.728$ with 11 degrees of freedom.

Thus, the differences between the distribution in column (3) and the "Student" distribution for $N = 5$ shown in column (6) are not only insignificant under the χ^2-test, but are so small as to be expected in a relatively small percentage of statistical experiments even if the "Student" z-distribution were the theoretically exact distribution of our ratios.

The usual moment coefficients of the distribution of observed z's in column (3), Table 5, are:

$$\mu_1' = 0.033533, \qquad \mu_3 = 0.254383, \qquad \beta_1 = 0.55955,$$

$$s = \sqrt{\mu_2} = 0.69799, \qquad \mu_4 = 2.22504, \qquad \beta_2 = 9.37353.$$

Since the value, 0.69799, of the standard deviation of the observed distribution differs very little from $1/\sqrt{N - 3} = 0.70711$, the normal curve fitted by using the standard deviation of the observed distribution (column 4, Table 5) differs very little from the normal curve with the origin at the population mean and standard deviation, $\sqrt{2}/2$, (column 5). Furthermore, the application of the χ^2-test to columns (4) and (5) of Table 5 with class frequencies in the "tails" grouped as above gives $\chi^2 = 2.91$ with 9 degrees of freedom.

The moment coefficients of the observed distribution indicate a markedly leptokurtic and somewhat skew distribution but the indications of skewness may be traced mainly and perhaps entirely to the presence of the two extreme variates at the upper end of the distribution and separated about three times the standard deviation from the next class frequency that differs from zero By

TABLE 5

Distribution of the ratios, $z = (\bar{y} - 7)/w$ in samples of $N = 5$ for Case I.

(1) $z = (\bar{y} - 7)/w$	(2) $t = z\sqrt{N-1} = 2z$	(3) Observed distribution	(4) Normal distribution fitted to observed column (3)	(5) Normal distribution of S.D $= \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{2}}$ in same units as z (measured from population mean)	(6) From the Student theoretical distribution for $N = 5$
−6.0	−12				0.1
−5.5	−11				0.1
−5 0	−10				0.1
−4.5	−9				0.2
−4.0	−8				0.3
−3.5	−7				0.6
−3.0	−6	2	0 05	0.1	1.3
−2.5	−5	1	0.75	0.4	2 7
−2 0	−4	5	3.6	6.2	7 0
−1.5	−3	17	27.7	32 0	21 0
−1.0	−2	67	98.5	105 9	70.5
−0 5	−1	216	210 8	217.2	217.5
0	0	357	279.3	275 4	356.2
0.5	1	226	225 7	217 2	217.5
1.0	2	75	111 7	105.9	70.5
1 5	3	22	33.7	32 0	21.0
2 0	4	5	6.2	6 2	7.0
2.5	5	1	0 75	0.4	2.7
3 0	6	3	0 05	0 1	1.3
3.5	7	0			0 6
4 0	8	0			0.3
4.5	9	0			0.2
5.0	10	2			0 1
5 5	11				0.1
6.0	12				0.1
		999	998.8	999.0	999.0

excluding these two variates from our calculations, we obtain the following moment coefficients:

$$\mu_1' = 0.023571, \qquad \mu_3 = 0.022264, \qquad \beta_1 = 0.0058786,$$

$$s = \sqrt{\mu_2} = 0.662202, \qquad \mu_4 = 1.009673. \qquad \beta_2 = 5.2507062.$$

In the observed distribution thus modified, by excluding the extreme upper class frequency 2, the evidence of skewness has disappeared.

Case II. For our Case II we have a frequency distribution as shown in Table 6.

TABLE 6

Totals in second throws of two dice-values of the stochastic variable.	2	3	4	5	6	7	8	9	10	11	12
Frequency . . .	1	4	9	16	25	36	35	32	27	20	11

Again, since with the uncorrected $\eta_{v\bar{u}}$, Table 6, we have $N\eta^2_{v\bar{u}} = 31.5$, and since $N\eta^2_{v\bar{u}}$ is approximately distributed as χ^2 with $a - 1 = 17$ degree of freedom, we have odds of the order of 100 to 1 against so large a value being a mere sampling fluctuation.

TABLE 7

Correlation of mean $\bar{u}$, and standard deviation $s_u = v$, of five items for Case II, mean of $\bar{u} = \bar{\bar{u}} = 6.971$. Correlation coefficient $r_{v\bar{u}} = -0.012 \pm 0.020$. $v = s_u = 2\,044$. Correlation ratio of v on $\bar{u}$, $\eta_{v\bar{u}} = 0.177$ (uncorrected).

v \ ū	37	41	45	49	53	57	61	65	69	73	77	81	85	89	93	97	101	105	109	f_v
39								1		1	1									3
37							1				3	1								5
35					1	2	1	4	4	2		2								16
33						1	5	4	3	5	2	2	1	1	1					25
31						4	6	9	6	4	6	1	1	1						38
29					3		8	10	16	8	8	5	5		2					65
27				1	2	4	10	7	7	17	13	4	1	2	1					69
25			1	1	2	10	10	17	8	19	11	5	4	3						91
23		1		3	5	5	11	11	14	13	10	7	10	3		1				94
21		1	4		3	15	21	16	13	16	13	9	5	3						118
19	1		1	1	12	12	7	12	9	16	19	7	13	5						115
17		1	1	6	5	11	14	12	16	17	20	10	7	5	2	3				130
15				1	7		14	8	8	3	6	5	4	3	3	1	1			64
13			1	1	7	10	3	11	3	13	4	1	6	1	1	2				64
11		1		2	3	4	5	7	9	9	4	4	5		3	1			1	58
09					2	1	2	4	4	6	2	3	1	3	1	1				30
07						2	3		1		2	2								10
05								2	1	2										5
$f_{\bar{u}}$	1	4	8	16	52	81	121	135	122	150	124	68	63	30	14	6	1	0	1	1000

Now proceeding to the analysis of variance, we substitute our numerical values derived from Table 7 in the identity

$$(4) \qquad \sum_{i=1}^{N} (v_i - \bar{v})^2 = \sum_{p=1}^{a} n_p(\bar{v}_p - \bar{v})^2 + \sum\sum (v_i - \bar{v}_p)^2$$

and obtain, in terms of class intervals as units,

$$10{,}871 = 340 + 10{,}531.$$

An outline of the analysis is exhibited in Table 8

TABLE 8

Variance	Degrees of freedom	Sum of squares	Mean square	z-test
Between arrays	17	340	20.00	$\frac{1}{2}\log_e 20.00 = 1.50$
Within arrays...	982	10,531	10.72	$\frac{1}{2}\log_e 10.72 = 1.18$
Total	999	10,871		Diff. = 0 32

The moment coefficients and β's which characterize the distribution in Table 6 are:

$$\text{Mean} = 7.972, \quad \mu_2 = 4.888, \quad \mu_3 = -1.755, \quad \mu_4 = 58.724,$$
$$\beta_1 = 0.0264, \quad \beta_2 = 2.449.$$

As in the linear case, samples of 5000 pairs of variates were drawn in sets of five by Miss Chesire. Analogous to Case I, our first concern is with the regression of the standard deviation, $s_u = v$, of u from a sample of five on its mean value, $\bar{u}$.

The correlation table for values of $\bar{u}$ and v is shown in Table 7. The correlation coefficient is
$r_{v\bar{u}} = -0.012 = \pm 0.021$, but the uncorrected correlation ratio of v on $\bar{u}$ is given by

$$\eta_{v\bar{u}} = 0.177.$$

After applying the Pearson and Student corrections, we obtain the corrected

$$\eta_{v\bar{u}} = 0.131.$$

When the sum of squares, 340, associated with variance between arrays is further analyzed into a part which could be represented by linear regression, and a part which represents deviations of the calculated means of arrays of v's from a straight regression line of v on $\bar{u}$, the deviations being measured parallel to the v-axis, we find that the part of the amount 340 represented by linear regression, would be only $Nr^2 s_v^2 = 1000\ (.000144)(10\ 871) = 1.6$.

Since both $r_{v\bar{u}} = -0.012 \pm 0.021$ and the small value, 1.6, as part of the sum of squares 340, may well be regarded as sampling fluctuations, we revert to the figures of Table 8.

The difference of the logarithms in the last column of Table 8, is 0.32, which corresponds to a level of significance of the general order of 100 to 1. Next, we calculate and plot on Table 7 the means of arrays of v's to give a general notion of the regression of v on $\bar{u}$. The location of these means of arrays suggests rather strongly that the regression of v on $\bar{u}$ is parabolic with the curvature concave downward as we should expect from the fact that $\beta_2 - \beta_1 - 3 < 0$, where the β's pertain to the parent distribution

Next, we computed 1000 values of the "Student" ratio, $z = (u - 7.972)/v$,

for Case II. One of these ratios was infinite. A frequency distribution of the 999 determinate ratios is shown in column 3, Table 9.

The observed distribution (column 3) and the "Student" distribution (column 6) of Table 9, to be expected in samples of $N = 5$, when samples are drawn from a normal distribution, are in close agreement. In fact, when we group together the tail frequencies of the theoretical distribution until each of them is not less than 5, the result of testing the goodness of fit gives $\chi^2 = 17.187$ with 11 degrees of freedom. This gives a value in the neighborhood of 0.1 for the probability, P, that as large or larger deviations than that experienced will occur, due to chance fluctuations, in a single repetition of the experiment. In other words, on the basis of this test, the indications are that we should have in the long run, as large or larger deviations than we have experienced in this case, in about 10 per cent of a large number of sets of sampling of 1000 per set even when the sampling is from a normal distribution.

TABLE 9

Distribution of the ratio, $(\bar{u} - 7.972)/v$ *in samples of five for Case* II.

(1) $z = (\bar{u} - 7.972)/v$	(2) $t = z\sqrt{N-1} = 2z$	(3) Observed	(4) Normal distribution fitted to observed, Column (3).	(5) Normal distribution with S.D. $= \frac{1}{\sqrt{N-3}}$ and origin at population mean	(6) Student's z-distribution for normal parent population with $N = 5$
−5.5	−11	1			0.1
−5.0	−10				0.1
−4.5	−9				0.2
−4.0	−8				0 3
−3.5	−7				0 6
−3.0	−6		0.1	0.1	1.3
−2.5	−5	2	0.4	0.4	2 7
−2 0	−4	3	4 3	6 2	7 0
−1.5	−3	23	25.4	32.0	21 0
−1.0	−2	48	92.0	105 9	70.5
−0.5	−1	203	205.3	217 2	217.5
−0 0	0	380	278.4	275.4	356 2
0.5	1	226	231.4	217.2	217.5
1 0	2	72	117 5	105 9	70 5
1.5	3	24	36.5	32 0	21 0
2.0	4	9	6.9	6.2	7.0
2.5	5	3	0.8	0 4	2 7
3.0	6	4	0.1	0.1	1.3
3.5	7	1			6
4 0	8				3
4 5	9				.2
					.1
					1
					.1
Total		999	999.1	999 0	999 0
∞	∞	1			

SUMMARY

1. The linear correlation coefficient, r, of the mean and standard deviation differs insignificantly from 0 in each case.

2. The correlation ratio of the standard deviation on the mean differs significantly from 0, and the regression of the standard deviation on the mean conforms, in its general aspects, to expectation under the theory of Neyman [12].

3. The indeterminate "Student" ratio of the form, $\frac{0}{0}$, in Case I and that of the form, (constant)/0, in Case II are probably due in part to grouping into class intervals, but the infinite ratio would undoubtedly have had such a large value that it would be excluded from calculations under any one of the known criteria for rejection of extreme observations.

4. Although the rejection of one indeterminate ratio in each of the two cases is slightly disturbing, the evidence presented by our analysis of the experimental sampling lends support to the view that the results of the "Student" theory are almost certainly applicable, for many purposes, when the parent distributions are of such non-normal types as are involved in our sampling.

REFERENCES

[1] W. A. Shewhart and F. W. Winters, "Small samples—new experimental results," *Journal American Statistical Association*, Vol. 23 (1928), pp. 144–153.

[2] P. R. Rider, "On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes," *Biometrika*, Vol. 21 (1929), pp. 124–143. "On small samples from certain non-normal universes," *Annals of Mathematical Statistics*, Vol. 2, (1931), pp. 48–65.

[3] E. S. Pearson, "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations," *Biometrika*, Vol. 21 (1929), pp. 259–286.

[4] M. S. Bartlett, "The effect of non-normality on the t distribution," *Proceedings of the Cambridge Philosophical Society*, Vol. 31 (1935), pp. 223–231.

[5] R. C. Geary, "The distribution of "Student's" ratio for non-normal samples," *Supplement to the Journal of the Royal Statistical Society*, No. 2, 1936, pp. 178–184.

[6] H. L. Rietz, "Urn schemeta as a basis for the development of correlation theory," *Annals of Mathematics*, Vol. 21 (1920), pp. 306–322.

[7] E. S. Pearson, Leona Chesire, and Elena M. Oldis," Further experiments on the sampling distribution of the correlation coefficient," *Journal American Statistical Association*, Vol. 27 (1932), pp. 121–128.

[8] Karl Pearson, "On a correction to be made in the correlation ratio," *Biometrika*, Vol. 8, (1911–12), pp. 254–6.

[9] "Student," "The correction to be made in the correlation ratio for grouping," *Biometrika*, (1913), pp. 316–20.

[10] R. A. Fisher, *Statistical methods for research workers*, Fourth Edition, p. 237.

[11] J. Neyman, "On the correlation of mean and variance in samples drawn from an 'infinite' population," *Biometrika*, Vol. 18 (1926), pp. 401–13.

[12] H. L. Rietz, *Mathematical Statistics* (Carus Monograph), 1926, p. 91.

UNIVERSITY OF IOWA, IOWA CITY, IOWA.

THE PROBLEM OF m RANKINGS

By M. G Kendall and B. Babington Smith

1. **Introduction.** If n objects are ranked by m persons according to some quality of the objects there arises the problem: does the set of m rankings of n show any evidence of community of judgment among the m individuals? For example, if a number of pieces of poetry are ranked by students in order of preference, do the rankings support the supposition that the students have poetical tastes in common, and if so is there any strong degree of unanimity or only a faint degree?

The problem in its full generality permits of no assumption about the nature of the quality according to which the objects are ranked, other than that ranking is possible. No hypothesis is made that the quality is measurable, still less that there is some underlying frequency distribution to the quantiles of which the rankings correspond. The quality is to be thought of as linear in the sense that any two objects possessing it are either coincident or may be put in the relation "before and after." A metric may, of course, be imposed on this linear space by convention; but the relationship between objects is invariant under any transformation which stretches the scale of measurement. In particular, it is not a condition of the problem that the ranking shall be based on a distribution according to a normal variate.

It is permissible to denote the rankings by the *ordinal* numbers $1, 2, \cdots n$; but it is not permissible, without further discussion, to operate on these numbers as if they were cardinals. This point seems to have been inadequately appreciated. For instance, when $m = 2$ we have the familiar case of rank correlation between a pair of rankings; and this is frequently treated by subtracting corresponding ranks, squaring, and forming the Spearman coefficient

$$\rho = 1 - \frac{6S(d^2)}{n^3 - n}. \tag{1}$$

To justify this procedure it is necessary to explain what is meant, for example, by such a process as (4th minus 8th), or what the square of this difference of ordinal numbers represents.

It is worth stressing that the necessary transition from ordinals to cardinals can be made without invoking a scale of measurement. When we rank an object as first we mean, in effect, that no member of the set of n is preferred to it; when we rank it as the rth we mean that $(r - 1)$ objects are preferred to it. The ordinals of the ranking are then biunivocally related to the cardinals expressing the number of objects which are preferred. It is thus legitimate

to apply the laws of cardinal arithmetic to them. For example, if an object A is ranked r_1 by Brown, r_2 by Jones and r_3 by Robinson we may form the sum $(r_1 + r_2 + r_3)$, which is to be interpreted as meaning that, taking the preferences of the three persons together, there were $(r_1 + r_2 + r_3 - 3)$ cases in which some other object was preferred to A. The point is of some importance, in view of the prevailing practice of replacing ranks by quantiles of the normal distribution—a practice which cannot always be regarded as justifiable and is sometimes little short of desperate.

To fix the ideas, consider the following three rankings of six objects

Object:	A	B	C	D	E	F
	5	4	1	6	3	2
	2	3	1	5	6	4
	4	1	6	3	2	5
Sum of ranks	11	8	8	14	11	11

We may sum the ranks for each object, as shown. These sums (which must add to 63, and in general to $mn(n + 1)/2$) reflect the degree of resemblance among the rankings. If the resemblance were perfect the sums would be 3, 6, 9, 12, 15, 18 (though not necessarily, of course in that order) and in such a case would be as different as possible. On the other hand, when there is little or no resemblance, as in the example given, the sums are approximately equal. It is thus natural to take the variance of these sums as providing some measure of the concordance in the rankings. If S is the observed sum of squares of the deviations of sums of ranks from the mean value $m(n + 1)/2$ (i.e. is n times the variance) we may write

$$W = \frac{12S}{m^2(n^3 - n)} \tag{2}$$

and call W the coefficient of concordance. Here $m^2(n^3 - n)/12$ is the maximum possible value of S, occurring if there is complete unanimity in the rankings, so that W may vary from 0 to 1. In the example given, $S = 25.5$, $W = 0.16$.

The coefficient W has arisen in several ways.

(a) W is simply related to the average of the $\binom{m}{2}$ Spearman rank correlation coefficients between pairs of the m rankings. It is easy to show that the average ρ is given by

$$\rho_{av} = \frac{\dfrac{12S}{n^3 - n} - m}{m^2 - m} \tag{3}$$

$$= \frac{mW - 1}{m - 1} \tag{4}$$

ρ_{av} has been considered by Kelley [3] as a measure of average intercorrelation in rankings, but he gives no results for testing the significance of observed values.

It is to be noted that whereas W may vary from 0 to 1, ρ_{av} may vary from $-1/(m-1)$ to 1, i.e it is asymmetrical like the coefficient of intraclass correlation, to which it bears some resemblance.[1]

(b) Friedman [1] has considered a quantity χ_r^2 related to W by the equation

$$\chi_r^2 = m(n-1)W. \tag{5}$$

(c) Welch [6] and Pitman [5] have considered the problem of the distribution of variance in an array

$$a_1, a_2, \cdots a_n$$

$$b_1, b_2, \cdots b_n$$

etc., for permutations of the numbers a, b, etc. in rows.

This is more general than the ranking case, in which $a_1 \cdots a_n$, $b_1 \cdots b_n$ etc. reduce to permutations of the numbers $1 \cdots n$. Since S', the total sum of squares in an array of m rankings of n, is $m^2(n^3-n)/12$, we have

$$W = \frac{S}{S'} \tag{6}$$

i.e. the ratio of variance between columns to the total variance.

2. **Significance of W.** To test whether an observed value of W is significant it is necessary to consider the distribution of W (or, more conveniently, of S) in the universe observed by permuting the n ranks in all possible ways. No generality is lost by supposing one ranking fixed, and the others will then give rise to $(n!)^{m-1}$ values of S.

The actual distribution of W (or S), as will be seen below, is irregular for low values of m and n, and likely to be quite irregular for moderate values. It may, however, be shown that the first four moments of W are

$$\mu_1' \text{ (about 0)} = \frac{1}{m} \tag{7}$$

$$\mu_2 = \frac{2(m-1)}{m^3(n-1)} \tag{8}$$

$$\mu_3 = \frac{8(m-1)(m-2)}{m^5(n-1)^2} \tag{9}$$

$$\mu_4 = \frac{24(m-1)}{m^7(n-1)^2}\left\{\frac{25n^3-38n^2-35n+72}{25(n^3-n)} + 2(n-1)(m-2) + \frac{n+3}{2}(m-2)(m-3)\right\}. \tag{10}$$

[1] The Spearman rank correlation coefficient is the product-moment coefficient of correlation between the ranks considered as ordinary variate values. ρ_{av} is the intraclass correlation coefficient for the m sets of ranks, also considered as variate values.

Results equivalent to these for the first three moments were given by Friedman [1]; and for the first four moments by Pitman [5]

In a valuable contribution to the subject Friedman showed that the distribution of χ_r^2 tends to that of χ^2 with $(n-1)$ degrees of freedom as m tends to infinity and suggested the use of χ_r^2 (equation (5)) for an ordinary test of significance in the χ^2 distribution. This is satisfactory for moderately large values, but for small values it is subject to the disadvantage inherent in any attempt to represent a distribution of finite range by one of infinite range—the fit near the tails is not likely to be very good. An improvement is obtained by noting that the first four moments of the Type I distribution,

$$df = \frac{1}{B(p,q)} W^{p-1}(1-W)^{q-1} \tag{11}$$

are approximately those of W if m and n are moderately large, and

$$p = \frac{n-1}{2} - \frac{1}{m} \tag{12}$$

$$q = (m-1)\left\{\frac{n-1}{2} - \frac{1}{m}\right\}. \tag{13}$$

For practical purposes it is most convenient to put

$$z = \tfrac{1}{2}\log_e \frac{(m-1)W}{1-W} \tag{14}$$

so that z can be tested in Fisher's distribution with $(n-1) - \frac{2}{m}$ $(= n_1)$ and $(m-1)\left\{(n-1) - \frac{2}{m}\right\}$ $(= n_2)$ degrees of freedom.

There can be little doubt that this test is quite reliable for moderate values of m and n; but it has hitherto been far from clear how reliable it is for low values of m and n. This point we attempt to clear up in the present paper.

3. **Distribution of S.** For the case $m = 2$ the distribution of S is the same as the distribution of the $S(d^2)$ used in calculating Spearman's rank correlation coefficient. A table showing the distribution up to and including $n = 8$ has already been given (Kendall and others, [4]). Tables giving probabilities that specified values of χ_r^2 would be attained or exceeded were given by Friedman [1] for $n = 3$, $m = 2$–9; and $n = 4$, $m = 2$–4. We have taken this work somewhat further and obtained the distributions for $n = 3$, $m = 2$–10; $n = 4$, $m = 2$–6; and $n = 5$, $m = 3$. Tables 1–4 give the probabilities based on these distributions.

These distributions were obtained by two methods. The first consisted of building up the distribution for $(m+1)$ and n from that of m and n For

TABLE 1

Probability that a given value of S will be attained or exceeded, for $n = 3$ and values of m from 2 to 10

Values of m

S	2	3	4	5	6	7	8	9	10
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	.866	.944	.931	.954	.956	.964	.967	.971	974
6	.500	.528	.653	.691	.740	.768	.794	.814	.830
8	.167	.361	.431	.522	.570	.620	.654	.685	.710
14		.194	.273	.367	.430	.486	.531	.569	.601
18		.028	.125	.182	.252	.305	.355	.398	.436
24			.069	.124	.184	.237	.285	.328	.368
26			.042	.093	.142	.192	.236	.278	.316
32			.0046	.039	.072	.112	.149	.187	.222
38				.024	.052	.085	.120	.154	.187
42				.0085	.029	.051	.079	.107	.135
50				$.0^377$	.012	027	.047	.069	.092
54					.0081	.021	.038	.057	.078
56					.0055	.016	.030	.048	.066
62					.0017	.0084	.018	.031	.046
72					$.0^313$	.0036	.0099	.019	.030
74						.0027	.0080	.016	.026
78						.0012	.0048	.010	.018
86						$.0^332$	.0024	.0060	.012
96						$.0^332$	.0011	.0035	.0075
98						$.0^421$	$.0^386$	.0029	.0063
104							$.0^326$	.0013	.0034
114							$.0^461$	$.0^366$	.0020
122							$.0^461$	$.0^335$	.0013
126							$.0^461$	$.0^320$	$.0^383$
128							$.0^536$	$.0^497$	$.0^351$
134								$.0^454$	$.0^337$
146								$.0^411$	$.0^318$
150								$.0^411$	$.0^311$
152								$.0^411$	$.0^485$
158								$.0^411$	$.0^444$
162								0^660	$.0^420$
168									$.0^411$
182									$.0^521$
200									$.0^799$

TABLE 2

Probability that a given value of S will be attained or exceeded for $n = 4$ and $m = 3$ and 5

S	$m = 3$	$m = 5$	S	$m = 5$
1	1.000	1.000	61	.055
3	.958	.975	65	.044
5	.910	.944	67	.034
9	.727	.857	69	.031
11	.608	.771	73	.023
13	.524	.709	75	.020
17	.446	.652	77	.017
19	.342	.561	81	.012
21	.300	.521	83	.0087
25	.207	.445	85	.0067
27	.175	.408	89	.0055
29	.148	.372	91	.0031
33	.075	.298	93	.0023
35	.054	.260	97	.0018
37	.033	.226	99	.0016
41	.017	.210	101	.0014
43	.0017	.162	105	$.0^{3}64$
45	.0017	.141	107	$.0^{3}33$
49		.123	109	$.0^{3}21$
51		.107	113	$.0^{3}14$
53		.093	117	$.0^{4}48$
57		.075	125	$.0^{5}30$
59		.067		

example, with $m = 2$ and $n = 3$ we have the following values of the sums of ranks, measured about their mean:

Type			Frequency
−2	0	2	1
−2	1	1	2
−1	0	1	2
0	0	0	1

Here −2, 1, 1, and 2, −1, −1 are taken to be identical types, for they give the same value of S and will also give similar types when we proceed to $m = 3$ as follows.

In the case $m = 3$ each of the above type will appear added to the six permutations of −1, 0, 1; e.g. the type −2, 0, 2 will give one each of −3, 0, 3; −3, 1, 2;

TABLE 3

Probability that a given value of S will be attained or exceeded for $n = 4$ and $m = 2, 4$ and 6

S	$m = 2$	$m = 4$	$m = 6$	S	$m = 6$
0	1.000	1.000	1.000	82	.035
2	.958	.992	.996	84	.032
4	.833	.928	.957	86	.029
6	.792	.900	.940	88	.023
8	.625	.800	.874	90	.022
10	.542	.754	.844	94	.017
12	.458	.677	.789	96	.014
14	.375	.649	.772	98	.013
16	.208	.524	.679	100	.010
18	.167	.508	.668	102	.0096
20	.042	.432	.609	104	.0085
22		.389	.574	106	.0073
24		.355	.541	108	.0061
26		.324	.512	110	.0057
30		.242	.431	114	.0040
32		.200	.386	116	.0033
34		.190	.375	118	.0028
36		.158	.338	120	.0023
38		.141	.317	122	.0020
40		.105	.270	126	.0015
42		.094	.256	128	$.0^390$
44		.077	.230	130	$.0^387$
46		.068	.218	132	$.0^373$
48		.054	.197	134	$.0^365$
50		.052	.194	136	$.0^340$
52		.036	.163	138	$.0^336$
54		.033	.155	140	$.0^328$
56		.019	.127	144	$.0^324$
58		.014	.114	146	$.0^322$
62		.012	.108	148	$.0^312$
64		.0069	.089	150	$.0^495$
66		.0062	.088	152	$.0^462$
68		.0027	.073	154	$.0^446$
70		.0027	.066	158	0^424
72		.0016	.060	160	$.0^416$
74		$.0^394$	.056	162	$.0^412$
76		$.0^394$	.043	164	$.0^580$
78		$.0^394$	.041	170	$.0^524$
80		$.0^472$	.037	180	$.0^613$

TABLE 4

Probability that a given value of S will be attained or exceeded, for n = 5 and m = 3

S	m = 3	S	m = 3
0	1.000	44	.236
2	1.000	46	.213
4	.988	48	.172
6	.972	50	.163
8	.941	52	.127
10	.914	54	.117
12	.845	56	.096
14	.831	58	.080
16	.768	60	.063
18	.720	62	.056
20	.682	64	.045
22	.649	66	.038
24	.595	68	.028
26	.559	70	.026
28	.493	72	.017
30	.475	74	.015
32	.432	76	.0078
34	.406	78	.0053
36	.347	80	.0040
38	.326	82	.0028
40	.291	86	$.0^{3}90$
42	.253	90	$.0^{4}69$

$-2, -1, 3$; $-2, 1, 1$; $-1, -1, 2$; and $-1, 0, 1$. These types are then counted for each of the four basic types of $m = 2$ and we get:

Type			Frequency
−3	0	3	1
−3	1	2	6
−2	0	2	6
−2	1	1	6
−1	0	1	15
0	0	0	2

The case $m = 4$ is treated by considering the numbers of types obtained by adding the six permutations of $-1, 0, 1$ to the types for $m = 3$; and so on.

This method is quite convenient for $n = 2$ and $n = 3$. For $n = 4$ it becomes difficult owing to the labour of considering 24 permutations at each stage and to the increase in the number of types. For $n = 5$ there are 120 permutations and the labour becomes excessive.

The second method employed is a generalisation of a procedure we used for the Spearman coefficient. Taking first of all the case $m = 2$, consider the array

$$\begin{array}{cccc} a^2 & a^3 & a^4 & \dots a^{(n+1)} \\ a^3 & a^4 & a^5 & \dots a^{(n+2)} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ a^{(n+1)} & a^{(n+2)} & a^{(n+3)} & \dots a^{2n} \end{array}$$

Any permissible set of values of the sums of ranks is obtained by selecting n entries from this array so that no entry appears more than once in the same row or column. If then, subtracting from each index the mean $(n + 1)$ and squaring, we write

$$E = \begin{Vmatrix} a^{(n-1)^2} & a^{(n-2)^2} & \dots & a^1 & a^0 \\ a^{(n-2)^2} & a^{(n-3)^2} & \dots & a^0 & a^1 \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ a^0 & a^1 & \dots & a^{(n-2)^2} & a^{(n-1)^2} \end{Vmatrix} \tag{15}$$

the values of S are the powers of a in E when it is expanded as a sum of $n!$ terms each of which is obtained by multiplying n factors which do not appear in the same row or column. The distribution of S is arrayed by the expansion of E, the number of values of any S being the coefficient of a^S in the expansion.

Similarly, for m rankings, the distribution of S is given by the expansion of an m- dimensional E-function. For example, with $m = 3$ there would be a three-dimensional E-function the bottom plane of which would be

$$\begin{array}{cccc} a^{\left\{3-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} \\ a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{5-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{2n+2-\frac{3(n+1)}{2}\right\}^2} \end{array}$$

The plane above this would be

$$\begin{array}{ccc} a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \dots\dots & \dots\dots & \dots\dots \\ a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{2n+3-\frac{3(n+1)}{2}\right\}^2} \end{array}$$

and so on.

The E-function is difficult to handle in more than three dimensions, but for the two and three dimensional case it is manageable and we used it to obtain the distribution of S for $n = 5$ and $m = 3$.

4. **Adequacy of the z-test.** Tables 1–4 provide exact tests for the values of m and n there given. It remains to be seen how good the ordinary z-test applied to W would be for higher values. It may be presumed that if the test is satis-

factory for any particular value of m and n for which exact results are available, it will be so for higher values of m and n. Since, for ordinary purposes the significance points of z as tabled by Fisher and Yates [2] would be employed, the most useful comparison would seem to be between those tables and the extreme values of tables 1–4.

For $n = 3$, $m = 9$, the 1% level is given approximately by $S = 78$. We have, testing for such a value, $W = 0.4814$, $z = 1.002$, $n_1 = \frac{16}{9}$, $n_2 = \frac{128}{9}$. By linear interpolation of reciprocals in the tables of z we find for the 1% point and these degrees of freedom $z = 0.954$. The correspondence is hardly satisfactory, and the z test might lead to incorrect inferences in practice. Matters improve a good deal, however, if we make continuity corrections, by subtracting unity from S before calculating W, and increasing by two the divisor $m^2(n^3 - n)/12$, so as to allow for the finite range. In this case $z = 0.979$.

For $n = 4$, $m = 6$ the 1% point is approximately $S = 100$. We have $W = 0.5556$, $z = 0\ 916$, $n_1 = 8/3$, $n_2 = 40/3$. By linear interpolation as before we find $z = 0.888$.

Continuity corrections again materially improve the agreement, giving a value of $z = 0.893$.

For $n = 5$ $m = 3$ there is no very convenient value of S close to the 1% point. For $P = 0.015$ $S = 74$ and for $P = 0.0078$ $S = 76$.

For $S = 74$ (with continuity corrections) $z = 1\ 020$
$S = 76$ (" " ") $z = 1\ 089$

By interpolation from the tables $z = 1.075$. The use of the z test would lead to the correct conclusion that a value of S equal to 74 falls below, and that of 76 above, the 1% point.

For values of m and n not included in Tables 1–4 it thus appears that the z-test with continuity corrections will give sufficiently accurate results, if n is greater than 3, at the 1% points. It may be presumed that the results at the 5% points are equally good and probably better. But for finer values of significance, such as 0.1%, it is doubtful whether the test is sound. The tails of the distribution of S for moderate values of m and n are very irregular.

For instance, the following is the tail of the distribution of S for $n = 3, m = 10$ (the total distribution being 10,077,696):

S	Frequency	S	Frequency
96	11,340	146	740
98	30,090	150	252
104	13,830	152	420
114	7,380	158	240
122	4,200	162	90
126	3,240	168	90
128	1,450	182	20
134	1,860	200	1

and the following is the tail for $n = 4$ $m = 6$ (the total being 7,962,624):

S	Frequency	S	Frequency	S	Frequency
100	5536	122	4100	146	810
102	8160	126	4480	148	225
104	10260	128	240	150	264
106	8850	130	1152	152	120
108	3920	132	660	154	180
110	13344	134	1980	158	60
114	5460	136	300	160	36
116	3870	138	600	162	30
118	3900	140	312	164	45
120	2472	144	100	170	18
				180	1

Irregularities of this kind run all through the distributions we have obtained, and frequency diagrams present the same sort of features we have noticed in the case $m = 2$ (Kendall and others, [4]). The representation of such distributions by continuous functions, no matter how close their lower moments, is obviously to be used with some care. Although the B-distribution or the associated z-distribution will give reasonable significance tests at levels of 1% or greater, they will probably be inadequate to represent frequencies occurring in narrow ranges.

5. **Some Experimental Distributions.** In some previous work we obtained a number of random permutations of the numbers 1–10 and 1–20. These were used to derive some experimental distributions of S which may be worth recording. Table 5 gives the distribution for 200 sets of pentads of 10 and Table 6 that for 100 triads of 20. In the distribution of Table 5, the mean of the grouped distribution is 404. The theoretical mean is 412.5 with a standard error of 12.3 In Table 6 the mean is 1936, the theoretical mean being 1995 with s.e. 53. The distributions accord quite well with expectation.

In conclusion we give two examples to illustrate some points of importance in ranking work. The first is a case in which ranks appear as the primary variate and in which the assumption of normality is clearly illegitimate.

6. **Example 1.** In some experiments in random series a pack of ordinary playing cards was shuffled and the order of the 13 cards of each suit from the top of the pack was noted. The pack was then reshuffled and again the orders noted. This was done 28 times. The question we wished to discuss was whether the shuffling was good, in the sense that the cards were thoroughly mixed at each shuffle.

Here, for each suit, say diamonds, we have 28 rankings of 13. The sums of ranks were 183, 137, 171, 207, 188, 160, 225, 174, 216, 192, 236, 239, 220. The mean is 196, and $S = 11522$, W (without continuity corrections, which are not

TABLE 5

Experimental Distribution of S in 200 sets ($m = 5$, $n = 10$)

S	Frequency
0–	1
50–	2
100–	7
150–	9
200–	21
250–	22
300–	24
350–	26
400–	20
450–	17
500–	12
550–	11
600–	10
650–	4
700–	5
750–	3
800–	3
...	.
1000–	2
..	..
1250–	1
Total	200

TABLE 6

Experimental Distribution of S in 100 sets ($m = 3$, $n = 20$)

S	Frequency
800–	4
1000–	8
1200–	8
1400–	6
1600–	12
1800–	15
2000–	20
2200–	12
2400–	6
2600–	5
2800–	0
3000–	3
3200–	0
3400–	1
Total	100

worth making for these values of m and n) = 0.08075, z (equation (14)) = 0.432. This falls just beyond the 1% point.

Similarly for the clubs W was found to be 0.0535; for the hearts, 0.0245; and for spades, 0.0342. None of these values is significant and we conclude that the randomisation introduced by the shuffling was good, at all events, so far as this test was concerned. It may be added that the shuffling was done with much more care than would be taken in an ordinary game of cards.

In psychological work there has sometimes been a confusion between the determination of a measure of agreement between subjects and that of an objective order based on experimental rankings. It may therefore be as well to point out that in its psychological applications the test of W is one of concordance between judgments. There may be quite a high measure of agreement about something which is incorrect.

7. **Example 2.** A number of students were given 12 photographs of persons unknown to them, and asked to rank them in what they judged from the photographs to be their intelligence. For 16 students the sums of ranks were

$$112, 94, 101, 84, 97, 75, 104, 84, 102, 146, 125, 124$$

The mean is 104. $S = 4472$, $W = 0.1222$. $z = 0.368$, and is barely significant, being between the 1% and the 5% points

For 111 students the sums were

$$818, 670, 908, 410, 706, 526, 780, 485, 596, 1044, 959, 756$$

$$W = 0.2378, z = 1.768$$

This is highly significant and it is to be inferred that community of judgment exists between students or groups of students. But there was little relationship between the judgments and the intelligence of the photographed subjects as given by the Binet Intelligence Quotient.

Note added in proof:

While this paper was passing through the press Professor W. Allen Wallis, of Stanford University, kindly drew our attention to some unpublished work of his own on this subject. Professor Wallis had also arrived at the coefficient W which, he points out, is the ranking analogue of the correlation ratio His paper is, we understand, on the point of publication.

REFERENCES

[1] M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Jour. Am. Stat. Assn.*, Vol 32(1937), p. 675.

[2] R A. Fisher, and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1938, Oliver & Boyd, Edinburgh.

[3] T. L Kelley, *Statistical Methods*, 1927.

[4] M. G Kendall, Sheila F. H Kendall, and Bernard Babington Smith, "The Distribution of Spearman's Coefficient of Rank Correlation in a Universe in which all Rankings Occur an Equal Number of Times," *Biometrika*, Vol. 30(1938), p 251.

[5] E. J. G Pitman, "Significance Tests which may be applied to Samples from any Populations: III The analysis of variance." *Biometrika*, Vol 29(1938), p. 322.

[6] B. L. Welch, "On the z- test in Randomised Blocks and Latin Squares," *Biometrika*, Vol. 29(1937), p. 21

LONDON, ENGLAND
AND
UNIVERSITY OF ST ANDREWS,
SCOTLAND.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

THE ALLOCATION OF SAMPLINGS AMONG SEVERAL STRATA

By J. Stevens Stock and Lester R. Frankel

1. **Introduction.** The problem of selecting a random sample so as to obtain optimum precision in making estimates has been the subject of inquiries by Bowley,[1] Neyman,[2] Sukhatme[3] and others. In estimating an average value of a variate in a population it is often profitable to stratify the universe into several homogeneous parts and sample at random within each of these parts. In order to obtain maximum efficiency for a given size of sample it appears that the number of samplings from each stratum should be proportional to the standard deviation of the characteristic under consideration and to the total number of units within the stratum. By distributing the sample in such a manner optimum precision will be obtained in estimating a general average.

However, it often happens that it is not the purpose of an investigation to study the aggregate of the universe. Evaluations and interrelations of characteristics in different groups or strata within the universe may be of importance. Thus, in cost-of-living surveys in a number of urban centers the object is to compare costs among the cities of different backgrounds. In such cases it is desirable for each city to have equal reliability so that each one may be treated as a unit. There are many other situations in the social sciences where analyses of this type are of importance.

2. **The Problem.** In general, the sampling problem is: Given several well defined areas of study and a fixed number of observations with which to make the survey, how best to distribute the observations such that each area will be represented with equal precision.

There are n observations to be distributed among m areas or strata. In the

[1] A. L. Bowley, "Measurement of the precision attained in sampling," *Bulletin de l'Institute International de Statistique* 1926 Rome, Tome XXII, 1-ere Livraison, 3-eme partie, pp. 1-62 (supplement).

[2] J. Neyman, "On the two different aspects of the representative method," *Journal of the Royal Statistical Society*, 1934, pp. 558-625.

[3] P. V. Sukhatme, "Contribution to the theory of the representative method," Supplement to the *Journal of the Royal Statistical Society*, Vol. II, 1935, pp. 253-268.

i-th stratum, if N_i is the total number of units, S_i^2 the variance of the characteristic to be measured, and n_i the size of the sample, the sampling error of the arithmetic mean is

$$\sigma_i = \sqrt{\frac{S_i^2}{n_i}\frac{(N_i - n_i)}{(N_i - 1)}}. \tag{1}$$

The problem then is, given N_i, numbers proportional to S_i^2 and n, to determine n_i such that

$$\sigma_1 = \sigma_2 = \cdots = \sigma_m.$$

3. **First Solution.** If we assume that the variances S_i^2 are all equal and that for $N_i - 1$ we may substitute N_i, we have

$$\frac{N_1 - n_1}{n_1 N_1} = \frac{N_2 - n_2}{n_2 N_2} = \cdots = \frac{N_m - n_m}{n_m N_m}. \tag{2}$$

From the total amount of money available and the cost per sampling unit we can determine the total number of observations to be apportioned among the m populations

$$n = \sum_1^m n_i. \tag{3}$$

We are able then to write m equations in m unknowns:
From (2) we may write $m - 1$ equations

$$\frac{1}{n_1} - \frac{1}{N_1} = \frac{1}{n_j} - \frac{1}{N_j} \qquad (j = 2, 3, \cdots m) \tag{4}$$

and from (3) we may write one equation.

$$n_1 + n_2 + \cdots + n_m = n. \tag{5}$$

But equations (4) are not easily soluble in their present form; they can be made linear by writing the approximation

$$\frac{1}{n_i} \equiv \frac{1}{L_i(1 + \alpha_i)} \doteq \frac{1 - \alpha_i}{i_i}.$$

Where L_i is some reasonable approximation of n_i chosen such that

$$\sum_1^m L_i = \sum_1^m n_i$$

and α_i is some small correction for L_i to be determined. We have then approximately,

$$\frac{1 - \alpha_1}{L_1} - \frac{1}{N_1} = \frac{1 - \alpha_j}{L_j} - \frac{1}{N_j} \qquad (j = 2, 3, \cdots m) \tag{6}$$

and from equation (5) we get

$$\alpha_1 L_1 + \alpha_2 L_2 + \cdots + \alpha_m L_m = 0. \tag{7}$$

If we write

$$\phi_i \equiv L_1 L_i \left(\frac{1}{N_1} - \frac{1}{N_i}\right) + L_1 - L_i$$

we may write (6) and (7) in the following form:

$$\begin{aligned}
&-L_2\alpha_1 + L_1\alpha_2 && = \phi_1 \\
&-L_3\alpha_1 \quad + L_1\alpha_3 && = \phi_3 \\
&\cdots\cdots\cdots\cdots\cdots\cdots && \\
&-L_m\alpha_1 \quad + L_m\alpha_1 && = \phi_m \\
&L_1\alpha_1 + L_2\alpha_2 + L_3\alpha_3 + \cdots + L_m\alpha_m && = 0
\end{aligned} \tag{8}$$

The matrix of the coefficients is

$$\begin{Vmatrix}
-L_2 & L_1 & 0 & \cdots & 0 \\
-L_3 & 0 & L_1 & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
-L_m & \cdot & \cdot & \cdots & L_1 \\
L_1 & L_2 & \cdot & \cdots & L_m
\end{Vmatrix} \tag{9}$$

From this matrix we find that

$$\alpha_1 = \frac{-\sum_2^m \phi_i L_i}{\sum_2^m L_i^2} \tag{10}$$

and from the general form of equation (8) we have

$$\alpha_i = \frac{\phi_i + L_i\alpha_1}{L_1} \tag{11}$$

These two equations (10) and (11) give us all the α_i. It is then only necessary to compute the second approximations of n_i by

$$L_i' = L_i(1 + \alpha_i) = n_i. \tag{12}$$

Closer approximations, though perhaps unnecessary, can be made by repeating the computation with the next approximations. The final approximations may be checked by substituting them in equations (4).

4. **Second Solution.** Sometimes the numbers S_i^2 are known or at least proportionate numbers can be estimated with a fair degree of accuracy for each area. We shall call these proportionate number ξ_i^2. We now have the conditions

$$\xi_1^2 \frac{N_1 - n_1}{n_1 N_1} = \xi_2^2 \frac{N_2 - n_2}{n_2 N_2} = \cdots = \xi_m^2 \frac{N_m - n_m}{n_m N_m} \tag{13}$$

and as before

$$\sum_1^m n_i = n .$$

We may write m equations in m unknowns, α_i, using the approximations L_i as before:[4]

$$\begin{aligned}
&- S_1^2 L_2 \alpha_1 + S_2^2 L_1 \alpha_2 && = \theta_2 \\
&- S_1^2 L_2 \alpha_1 \cdot \quad + S_3^2 L_1 \alpha_3 && = \theta_3 \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot && \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot && \\
&- S_1^2 L_m \alpha_1 \qquad\qquad + S_m^2 L_1 \alpha_m && = \theta_m \\
&L_1 \alpha_1 + L_2 \alpha_2 + \cdots + L_m \alpha_m && = 0
\end{aligned} \tag{14}$$

Where

$$\theta_i = L_1 S_i^2 - L_i S_i^2 + L_i L_1 \left(\frac{S_1^2}{N_1} - \frac{S_i^2}{N_i} \right). \tag{15}$$

Solving these m linear equations for α_i we get

$$\alpha_1 = \frac{-\sum_2^m \theta_i L_i / S_i^2}{S_1^2 \sum_2^m L_i^2 / S_i^2}$$

and from the general form of equations (14) we have

$$\alpha_i = \frac{\theta_i + S_1^2 L_i \alpha_1}{S_i^2 L_1} .$$

These α_i may be applied as before to the approximations L_i for new approximations L_i' of the numbers n_i.

5. **Remarks.** (i) In either case the applications of the corrections to the approximation L_i may be applied in two different ways:

$$L_i' = L_i(1 + \alpha_i) \tag{16}$$

$$L_i' = \frac{L_i}{1 - \alpha_i} . \tag{17}$$

When the corrections are applied according to (16) the sum of the new approximations adds up correctly to the total n, and no further adjustment need be made in the L_i' either for repeating the operation again for nearer approximations or for final use. If, however, the corrections are relatively large, say

[4] The numbers S_i^2 and ξ_i^2 may be used interchangeably since they are by hypothesis proportional.

greater than .10, there seems to be better convergence with the second approximations if formula (17) is used and the resulting L_i' adjusted proportionately such that they add up to n. These numbers then can again be adjusted with new α_i for final approximations.

(ii) The numbers S_i^2 or ξ_i^2 are not always estimable. If they are not known at all or are known to be all nearly equal the first solution is perhaps the more useful. If these numbers are known, and known to be different, the second solution is necessary However, some saving in computation by the second method may be effected if the approximations L_i are first adjusted by the first solution before being entered into the computation of the second solution.

(iii) Further accuracy, though perhaps unnecessary, may be attained in the second solution by substituting throughout $S_i'^2$ for S_i^2 where

$$S_i'^2 = \frac{N_i}{N_i - 1} S_i^2$$

This substitution eliminates any slight inaccuracies caused by substituting N_i for $N_i - 1$.

(iv) The initial approximations L_i may in almost every case be gotten from the following formula:

$$L_i = \frac{n}{m} - \left(\frac{n}{m}\right)^2 \left(\frac{1}{N_i} - \frac{1}{m} \sum_1^m \frac{1}{N_i}\right).$$

(v) In all that has been presented above it has been assumed that the sample has been drawn without replacements from a finite universe. Whether or not this assumption is tenable depends upon the particular object of the research.

6. **Example.** In the Survey of Youth in the Labor Market conducted by the Division of Research in the Works Progress Administration youth who completed the eighth grade in the school years 1928–1929, 1930–1931 and 1932–1933 were studied. In six cities, Duluth, Denver, Birmingham, Seattle, San Francisco, and St. Louis random samples from school records were selected. Funds permitted the use of 40,000 schedules.

From school records it was possible to determine the total number of eighth grade graduates in each city for the years in question. The problem arose then as to what would be the most efficient method of distributing these 40,000 schedules among the six cities in order to compare the problems of youth.

Assuming equal variances within cities, quotas were computed for each of the cities. From Table 1, summarizing the computations, it can be seen that the quotas fall somewhere between proportionate and equal frequencies. This last result would be expected if samplings had been made from infinite universes.

7. **Note.** In the social sciences interest centers in deriving relationships among the various strata where each stratum is considered as a single unit. In such cases equal precision is desired. However, if the object of research is

TABLE 1

City	8th grade graduates	Initial approximation	First correction term	First approximation	Second correction term	Quotas	Percent sampled
Duluth, Minn.	5,500	4,000	−.02968	3,881	−.00077	3,878	70.51
Birmingham, Ala.	9,000	5,500	+.06641	6,399	+.00148	5,343	59.37
Denver, Colo.	12,500	6,000	−.02690	5,352	−.00164	6,409	51 27
Seattle, Wash.	15,000	6,500	+.07525	6,989	+.00257	7,007	46.71
San Francisco, Cal.	21,000	8,000	+.01425	8,114	−.00341	8,086	38 50
St. Louis, Mo.	31,000	10,000	−.07349	9,265	+ 00129	9,277	29.93
Total	94,000	40,000		40,000		40,000	

simply to draw contrasts between any two strata we would seek to minimize the standard error of the difference,

$$\sigma_{\Delta,k} = \sqrt{S_j'^2\left(\frac{1}{n_j} - \frac{1}{N_j}\right) + S_k'^2\left(\frac{1}{n_k} - \frac{1}{N_k}\right)}$$

subject to the condition,

$$\sum_1^m n_i = n .$$

This leads to the result

$$\frac{S_j'}{n_j} = \frac{S_k'}{n_k} .$$

Thus, the number of samplings from each stratum is, for all practical purposes, proportional to the standard deviations, irrespective of the size of the various strata.

WASHINGTON, D. C.

ON THE COEFFICIENTS OF THE EXPANSION OF $X^{(n)}$

BY J. A. JOSEPH

Let us construct the following triangular arrangement of numbers:

$$\begin{array}{ccccccccccccc}
 & & & & & 1 & & & & & & & \\
 & & & & 1 & & 1 & & & & & & \\
 & & & 1 & & 3 & & 2 & & & & & \\
 & & 1 & & 6 & & 11 & & 6 & & & & \\
 & 1 & & 10 & & 35 & & 50 & & 24 & & & \\
 & \cdot & & \cdot & & \cdot & & \cdot & & \cdot & & \cdot & \\
1 & f_1(n-1) & & f_2(n-1) & & \cdot & & \cdot & & \cdot & f_{n-2}(n-1) & & f_{n-1}(n-1) \\
1 & f_1(n) & & f_2(n) & & \cdot & & \cdot & & \cdot & f_{n-1}(n) & & f_n(n)
\end{array}$$

where the n-th row can be constructed from the preceding row by means of the expression

$$n\, f_i(n-1) + f_{i+1}(n-1) = f_{i+1}(n). \tag{1}$$

For example, the element 35 in the middle of the 4th row is obtained from the two elements immediately above it, $4 \cdot 6 + 11 = 35$. (The top element is counted as the zeroth row.)

The elements in the $(n-1)$st row are the coefficients in the expansion of $x^{(n)}$ as a function of x, using the notation of the calculus of finite differences. For example,

$$\begin{aligned} x^{(4)} &= x(x-1)(x-2)(x-3) \\ &= x^4 - 6x^3 + 11x^2 - 6x. \end{aligned}$$

Of course, the signs of the coefficients alternate.

The function $f_i(n)$ is the sum of the products of the first n integers taken i at a time, namely

$$f_i(n) = \sum^{n} \epsilon_1 \epsilon_2 \cdots \epsilon_i \tag{2}$$

the summation being a symmetric function of the integers $1, 2, 3, \cdots, n$.

Equation (1) can be written as a linear, first order difference equation,

$$\begin{aligned} \Delta f_{i+1}(n-1) \equiv f_{i+1}(n) - f_{i+1}(n-1) &= n \cdot f_i(n-1) \\ f_{i+1}(n-1) &= \Delta^{-1}[n \cdot f_i(n-1)]. \end{aligned} \tag{3}$$

Since $f_0(n) = 1$ for all values of n, we can find $f_1(n)$, and consequently $f_2(n)$, and so on. Thus

$$\begin{aligned} f_1(n-1) &= \Delta^{-1} n = \frac{n^{(2)}}{2} \\ f_2(n-1) &= \Delta^{-1}\left[n \cdot \frac{n^{(2)}}{2}\right] = \frac{3n^{(4)} + 8n^{(3)}}{24} \\ f_3(n-1) &= \Delta^{-1}\left[n\left(\frac{3n^{(4)} + 8n^{(3)}}{24}\right)\right] \\ &= \frac{n^{(5)} + 8n^{(5)} + 12n^{(4)}}{48}. \end{aligned} \tag{4}$$

The following theorems are true for the "triangle":

THEOREM 1. *The sum of the elements in any n-th row is equal to $(n+1)!$, namely,*

$$\sum_{i=0}^{i} f_i(n) = (n+1)! \tag{5}$$

THEOREM 2. *The sum of the even elements of any row is equal to the sum of the odd elements, or*

(6) $$\sum_{i=0}^{n} (-1)^i f_i(n) = 0.$$

From these coefficients we can generate the Bernoulli numbers:

$$B_0 = \frac{1}{2}$$

$$B_0 - B_1 = \frac{2!}{3}$$

$$2B_0 - 3B_1 + B_2 = \frac{3!}{4}$$

(7) $$6B_0 - 11B_1 + 6B_2 - B_3 = \frac{4!}{5}$$

$$24B_0 - 50B_1 + 35B_2 - 10B_3 + B_4 = \frac{5!}{6}$$

$$\cdots\cdots\cdots\cdots$$

$$f_n(n)B_0 - f_{n-1}(n)B_1 + f_{n-2}(n)B_2 - \cdots + (-1)^n f_0 B_n = \frac{(n+1)!}{n+2}$$

Or, as a determinant:

(8) $$| B_n | = \begin{vmatrix} \frac{1}{2} & 1 & 0 & 0 & \cdots & 0 \\ \frac{2!}{3} & 1 & 1 & 0 & \cdots & 0 \\ \frac{3!}{4} & 2 & 3 & 1 & \cdots & 0 \\ \frac{4!}{5} & 6 & 11 & 6 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{(n+1)!}{n+2} & f_n(n) & f_{n-1}(n) & f_{n-2}(n) & \cdots & f_1(n) \end{vmatrix}$$

giving

$$B_0 = \tfrac{1}{2}, \quad B_1 = -\tfrac{1}{6}, \quad B_2 = B_4 = B_6 = \cdots = B_{2n} = 0,$$

$$B_3 = \tfrac{1}{30}, \quad B_5 = -\tfrac{1}{42}, \cdots .$$

We may now take another "triangle":

$$\begin{array}{ccccccccccccc} & & & & & & 1 & & & & & & \\ & & & & & 1 & & 1 & & & & & \\ & & & & 1 & & 3 & & 1 & & & & \\ & & & 1 & & 6 & & 7 & & 1 & & & \\ & & 1 & & 10 & & 25 & & 15 & & 1 & & \\ & & & \cdot & & \cdot & & \cdot & & \cdot & & \cdot & \\ & 1 & & F_1(n-1) & & F_2(n-1) & \cdots\cdots & F_{n-2}(n-1) & & F_{n-1}(n-1) & & & \\ 1 & & F_1(n) & & F_2(n) & & \cdots\cdots\cdots & & F_{n-1}(n) & & F_n(n) & & \end{array}$$

where the n-th row is obtained from the preceding row by the expression

$$(n - i)F_i(n - 1) + F_{i+1}(n - 1) = F_{i+1}(n). \tag{9}$$

For example, from the third row: 1, 6, 7, 1, we obtain the fourth row: $1, 4 \cdot 1 + 6 = 10, 3\ 6 + 7 = 25, 2 \cdot 7 + 1 = 15, 1$. The following theorem is true for the $F_i(n)$:

THEOREM 3 *The elements in the $(n - 1)$st row are the coefficients in the expansion of x^n as a function of the factorials $x^{(i)}$.*

For example:

$$x^4 = x^{(4)} + 6x^{(3)} + 7x^{(2)} + x.$$

From the generating equation (9) we can obtain, as before, the form of the functions $F_0(n)$, $F_1(n)$, $\cdots$

$$\begin{aligned} \Delta F_{i+1}(n - 1) &\equiv F_{i+1}(n) - F_{i+1}(n - 1) = (n - i)F_i(n - 1) \\ F_{i+1}(n - 1) &= \Delta^{-1}[(n - i)F_i(n - 1)]. \end{aligned} \tag{10}$$

Since $F_0(n) = 1$ for all n

$$\begin{aligned} F_1(n - 1) &= \Delta^{-1}n = \frac{n^{(2)}}{2} \\ F_2(n - 1) &= \Delta^{-1}\left[(n - 1)\frac{n^{(2)}}{2}\right] = \frac{3n^{(4)} + 4n^{(3)}}{24} \\ F_3(n - 1) &= \Delta^{-1}\left[(n - 2)\frac{3n^{(4)} + 4n^{(3)}}{24}\right] \\ &= \frac{n^{(6)} + 4n^{(5)} + 2n^{(4)}}{48}. \end{aligned} \tag{11}$$

From these coefficients we can generate the numbers of Laplace (the numbers L_m below must be divided by $m!$ to yield the numbers of Laplace):

$$\begin{aligned} L_1 &= \tfrac{1}{2} \\ L_1 + L_2 &= \tfrac{1}{3} \\ L_1 + 3L_2 + L_3 &= \tfrac{1}{4} \\ L_1 + 7L_2 + 6L_3 + L_4 &= \tfrac{1}{5} \\ L_1 + 15L_2 + 25L_3 + 10L_4 + L_5 &= \tfrac{1}{6} \\ &\cdots \\ L_1 + F_{n-1}(n)L_2 + F_{n-2}(n)L_3 + \cdots + L_{n-1} &= \frac{1}{n + 1} \end{aligned} \tag{12}$$

giving

$$L_1 = \tfrac{1}{2}, \quad L_2 = -\tfrac{1}{6}, \quad L_3 = \tfrac{1}{4}, \quad L_4 = -\tfrac{19}{30}, \quad L_5 = \tfrac{9}{4}$$

A determinantal solution is also obvious.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

ON THE PROBABILITY OF ATTAINING A GIVEN STANDARD DEVIATION RATIO IN AN INFINITE SERIES OF TRIALS

By Joseph A. Greenwood and T. N. E. Greville

Suppose an event with constant probability p of occurrence to be repeated an infinite number of times, and suppose the ratio of the deviation from the expected number of successes to the standard deviation $\sqrt{npq}$ to be recomputed after each trial. We are interested in the probability that this ratio will at some time equal or exceed some positive number k. It is not difficult to show that the value of this probability is unity, but as the fact has not, to our knowledge, been previously pointed out in the literature, we give the following proof.

Let x_n denote the number of successes obtained in the first n trials, let

$$t_n = \frac{x_n - np}{\sqrt{npq}},$$

and let P denote the probability that, for some n, $t_n \geqq k$. We shall prove that $P = 1$. To do this, let the infinite series of trials be subdivided into consecutive, mutually exclusive subseries of finite length, and let m_i denote the number of trials in the i-th subseries. Let $N_i = \sum_{j=1}^{i-1} m_j$ for $i \geqq 2$, while $N_1 = 0$. Let k' be any number greater than k, and let m_i be so chosen that

$$m_i \geqq \frac{k'^2 p}{q} \qquad \text{for every } i, \tag{1}$$

and

$$\sqrt{m_i}\left(k' - k\sqrt{\frac{N_i}{m_i} + 1}\right) \geqq N_i\sqrt{\frac{p}{q}} \qquad \text{for } i \geqq 2. \tag{2}$$

It follows from (1) that

$$m_i \geqq m_i p + k'\sqrt{m_i pq} \qquad \text{for every } i. \tag{3}$$

It follows from (2) that

$$m_i p + k'\sqrt{m_i pq} \geqq (N_i + m_i)p + k\sqrt{(N_i + m_i)pq} \qquad \text{for every } i. \tag{4}$$

Let y_i denote the number of successes in the i-th subseries. It is evident from (4) that if

$$y_i \geqq m_i p + k'\sqrt{m_i pq} \tag{5}$$

for any i, then

$$t_{N_i + m_i} \geqq k.$$

Hence P is at least equal to the probability that (5) holds for some i.

Let p_i denote the probability that (5) holds for a particular i. It follows from (3) that, for every i, $p_i > 0$. Moreover, there exists a positive integer M

and a number $h > 0$, such that if $m_i \geqq M$, $p_i \geqq h$.[1] Since there is but a finite number of possible values of m_i less than M, there is a number $p_0 > 0$ such that $p_i \geqq p_0$ for every i. Hence the probability that (5) holds for no value of i is at most

$$\lim_{s \to \infty} (1 - p_0)^s = 0.$$

Therefore, the probability that (5) holds for some i is unity.

DUKE UNIVERSITY
AND
UNIVERSITY OF MICHIGAN.

[1] Uspensky, J. V., *Introduction to Mathematical Probability*, p. 129.

CONTRIBUTIONS TO THE THEORY OF STATISTICAL ESTIMATION AND TESTING HYPOTHESES[1]

By Abraham Wald

1. **Introduction.** Let us consider a family of systems of n variates $X_1(\theta^{(1)}, \cdots, \theta^{(k)}), \cdots, X_n(\theta^{(1)}, \cdots, \theta^{(k)})$ depending on k parameters $\theta^{(1)}, \cdots, \theta^{(k)}$. A system of k values $\theta^{(1)}, \cdots, \theta^{(k)}$ can be represented in the k-dimensional parameter space by the point θ with the co-ordinates $\theta^{(1)}, \cdots, \theta^{(k)}$. Denote by Ω the set of all possible points θ. For any point θ of Ω we shall denote by $P(E \epsilon w|\theta)$ the probability that the sample point $E = (x_1, \cdots, x_n)$ falls into the region w of the n-dimensional sample space, where x_j denotes the observed value of the variate $X_j(\theta)(j = 1, \cdots, n)$. The distribution $P(E \epsilon w|\theta)$ is supposed to be known for any point θ of Ω. In the theory of testing hypotheses and of statistical estimation we have to deal with problems of the following type: A sample point $E = (x_1, \cdots, x_n)$ of the n-dimensional sample space is given. We know that x_j is the observed value of $X_j(\theta)$ but we do not know the parameter point θ, and we have to draw inferences about θ by means of the sample point observed. The assumption that θ belongs to a certain subset ω of Ω is called a hypothesis. We shall deal in this paper with the following general problem: Let us consider a system S of subsets of Ω. Denote by H_ω the hypothesis corresponding to the element ω of S, and by H_S the system of all hypotheses corresponding to all elements of S. We have to decide by means of the observed sample point E which hypothesis of the system H_S should be accepted. That is to say for each H_ω we have to determine a region of acceptance M_ω in the n-dimensional sample space. The hypothesis H_ω will be accepted if and only if the sample point E falls in the region M_ω. M_ω and $M_{\omega'}$ are disjoint if $\omega \neq \omega'$. The statistical problem is the question as to how the system M_S of all regions M_ω should be chosen.

The problem in this formulation is very general. It contains the problems of testing hypotheses and of statistical estimation treated in the literature.[2] For instance if we want to test the hypothesis H_ω corresponding to a certain subset ω of Ω, the system of hypotheses H_S consists only of the two hypotheses H_ω and $H_{\bar{\omega}}$ where $\bar{\omega}$ denotes the subset of Ω complementary to ω. If we want to estimate θ by a unique point, then S is the system of all points of Ω. In the theory of confidence intervals we estimate one of the parameter co-ordinates $\theta^{(1)}, \cdots, \theta^{(k)}$,

[1] Research under a grant-in-aid from the Carnegie Corporation of New York.

[2] See, for instance, J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Transactions of the Royal Society*, London, Vol. 231 (1937), pp. 333-380.

say $\theta^{(1)}$, by an interval. In this case S is a certain system of subsets ω of the following type: ω is the set of all points $\theta = (\theta^{(1)}, \cdots, \theta^{(k)})$ for which $\theta^{(1)}$ lies in a certain interval $[a, b]$. The problem in our formulation covers also cases which, as far as I know, have not yet been treated. Consider for instance 3 subsets ω_1, ω_2 and ω_3 of Ω such that the sum of them is equal to Ω. It may be that we are interested only to know in which of the subsets ω_1, ω_2, ω_3 the unknown parameter point lies. In this case the system of hypotheses H_S consists only of the 3 hypotheses H_{ω_1}, H_{ω_2} and H_{ω_3}. Cases like this might be of practical interest.

For the determination of the "best" system (in a certain sense) of regions of acceptance we shall use methods and principles which are closely related to those of the Neyman-Pearson theory of testing hypotheses. In the Neyman-Pearson theory two types of error are considered. Let $\theta = \theta_1$ be the hypothesis to be tested, where θ_1 denotes a certain point of the parameter space. Denote this hypothesis by H_1 and the hypothesis $\theta \neq \theta_1$ by $\bar{H}$. The type I error is that which is made by rejecting H_1 when it is true. The type II error is made by accepting H_1 when it is false. The fundamental principle in the Neyman-Pearson theory can be formulated as follows: among all critical regions (regions of rejection of H_1, i.e. regions of acceptance of $\bar{H}$) for which the probability of type I error is equal to a given constant α, we have to choose that region for which the probability of type II error is a minimum. The difficulty which arises here lies in the circumstance that the probability of type II error depends on the true parameter point θ. That is to say, if the critical region is given the probability of type II error will be a function of the true parameter point θ. Since we do not know the true parameter point θ, we want to have a critical region which minimizes the probability of type II error with respect to any possible alternative hypothesis $\theta = \theta_2 \neq \theta_1$. If such a common best critical region exists, then the problem is solved. But such cases are rather exceptional. If a common best critical region does not exist, Neyman and Pearson consider unbiased critical regions of different types,[3] which minimize the type II error locally, that is to say with respect to alternative hypotheses in the neighborhood of the hypothesis considered. In this paper we develop methods for the determination of a system of regions of acceptance taking in account type II errors also relative to alternative hypotheses not lying in the neighborhood of the hypothesis to be tested.

2. **Some Definitions.** Let us denote by Ω the set of all possible parameter points θ and by S a system of subsets of Ω. If ρ denotes the sum of the elements of a subset σ of S, then we shall denote ΣM_ω by M_ρ, where M_ω denotes the

[3] J. Neyman and E. S. Pearson: *Statistical Research Memoirs*, Volumes I and II. The authors consider also unbiased regions of type A_1 for which the probability of type II error with respect to every alternative hypothesis is not greater than for any other unbiased region of the same size. However regions of type A_1 do not always exist (the existence of such regions has been proved for a special but important class of cases).

region of acceptance of H_ω and the summation is to be taken over all elements ω of σ.

Definition 1. Denote by M_S and M_S' two different systems of regions of acceptance corresponding to the same system H_S of hypotheses. The systems M_S and M_S' are said to be equivalent if for each point θ of Ω and for every ρ which is a sum of elements of S which does not contain θ, the equation

$$P(E \in M_\rho' \mid \theta) = P(E \in M_\rho \mid \theta)$$

holds, where M_ρ' denotes the region according to the system M_S' and M_ρ denotes the region according to the system M_S.

Definition 2. Denote by M_S and M_S' two different systems of regions of acceptance corresponding to the same system of hypotheses. The system M_S' is said to be absolutely better than the system M_S if they are not equivalent and if for each θ and for every ρ which is a sum of elements of S which does not contain θ the inequality

$$P(E \in M_\rho' \mid \theta) \leq P(E \in M_\rho \mid \theta)$$

holds.

Definition 3. A system M_S of regions of acceptance is said to be admissible if no absolutely better system of regions exists.

3. **The problem of the choice of M_S.** The choice of M_S will in general be affected by the following two circumstances:

(1) We do not attribute the same importance to each error. For instance the acceptance of the hypothesis that θ lies in a certain interval I has in general more serious consequences if θ is far from I than if θ is near to I. The choice of M_S will in general depend on the relative importance of the different possible errors.

(2) In some cases we have a priori more confidence that the true parameter point lies in a certain interval I than in some other cases. The choice of M_S will in general be affected also by this fact. Let us illustrate this by an example We have two coins, a new and an old one and we want to test for both coins whether the probability p of tossing head is equal to $\frac{1}{2}$. Let us assume that we make 100 tosses with each of the coins and we get head 40 times in each case. Since we have a priori no very great confidence that the old coin is unbiased, the fact that head occured only 40 times will suffice to reject the hypothesis that for the old coin $p = \frac{1}{2}$. But in the case of the new coin, having much greater a priori confidence that it is unbiased, we shall perhaps not reject the hypothesis $p = \frac{1}{2}$ and we shall rather assume that a somewhat improbable event occurred. That is to say, we do not choose the same critical region in both cases due to the fact that our a priori confidence for $p = \frac{1}{2}$ is in the case of the new coin greater than in the case of the old one.

In order to study the dependence of the choice of M_S on the two circumstances

mentioned, let us introduce a weight function for the possible errors and an a priori probability distribution for the unknown parameter θ. The weight function $W(\theta, \omega)$ is a real valued non-negative function defined for all points θ of Ω and for all elements ω of S, which expresses the relative importance of the error committed by accepting H_ω when θ is true. If θ is contained in ω, $W(\theta, \omega)$ is, of course, equal to zero. The question as to how the form of the weight function $W(\theta, \omega)$ should be determined, is not a mathematical or statistical one. The statistician who wants to test certain hypotheses must first determine the relative importance of all possible errors which will entirely depend on the special purposes of his investigation. If that is done, we shall in general be able to give a more satisfactory answer to the question as to how the system of regions of acceptance should be chosen. In many cases, especially in statistical questions concerning industrial production, we are able to express the importance of an error in terms of money, that is to say, we can express the loss caused by the error considered in terms of money. We shall also say that $W(\theta, \omega)$ is the loss caused by accepting H_ω when θ is true.

The situation regarding the introduction of an a priori probability distribution of θ is entirely different. First, the objection can be made against it, as Neyman has pointed out, that θ is merely an unknown constant and not a variate, hence it makes no sense to speak of the probability distribution of θ. Second, even if we may assume that θ is a variate, we have in general no possibility of determining the distribution of θ and any assumptions regarding this distribution are of hypothetical character. On account of these facts the determination of the system of regions of acceptance should be independent of any a priori probability considerations. The "best" system of regions of acceptance, which we shall define later, will depend only on the weight function of the errors. The reason why we introduce here a hypothetical probability distribution of θ is simply that it proves to be useful in deducing certain theorems and in the calculation of the best system of regions of acceptance.

Let us denote by $f(\theta)$ a distribution function of θ. For the sake of simplicity let us assume that the probability density of the distribution $P(E \in w \mid \theta)$ exists in any point E of the sample space for any θ and denote it by $p(E \mid \theta)$. The expected value of the loss is given by

$$I = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE \tag{1}$$

where ω_E denotes the element of S corresponding to E (that is to say, ω_E is that element of S for which E is a point of the region of acceptance M_{ω_E}), and the integral is to be taken over the product of the sample space M with the parameter space Ω. The expected value I of the loss depends on the system M_S of regions of acceptance. The system M_S for which I becomes a minimum, can be regarded as the best system of regions relative to the given weight function and to the given a priori distribution of θ.

One can easily show the following: If M_S' is an absolutely better system of regions (in sense of the definition 2) than the system M_S, then for any weight

function $w(\theta, \omega)$ and for any a priori distribution $f(\theta)$ the expected value I' of the loss corresponding to M_S' is less than the expected value I of the loss corresponding to M_S. (For some exceptional weight and a priori distribution functions I' may be equal to I.)

Hence we can give the following rule: *We have to choose an admissible system of regions of acceptance.*

Now let us consider the question whether besides admissibility further restrictions upon the choice of M_S can be made. In order to see this, let us consider two admissible systems of regions M_S and M_S' which are not equivalent. One can easily show that there exist two weight functions $W_1(\theta, \omega)$, $W_2(\theta, \omega)$ and two a priori distributions $f_1(\theta)$ and $f_2(\theta)$ such that for $W_1(\theta, \omega)$ and $f_1(\theta)$ the expected value of the loss corresponding to M_S is less than that corresponding to M_S', and for $W_2(\theta, \omega)$ and $f_2(\theta)$ the expected value of the loss corresponding to M_S is greater than that corresponding to M_S'. Hence no absolute criteria can be given as to which of the systems M_S and M_S' should be chosen. In order to be able to make further restrictions upon the choice of M_S, we have to make assumptions regarding the form of the weight function. We shall deal with this question in section 6

4. **Calculation of admissible systems of regions.** As we have seen, we have to choose an admissible system of regions The question arises as to how we can find admissible systems of regions.

Provided that $p(E \mid \theta)$ is continuous in E and θ jointly, one can easily show that M_S' is an admissible system of regions if there exists a bounded, uniformly continuous and everywhere positive (except if θ is contained in ω) weight function $W(\theta, \omega)$ and an a priori distribution $f(\theta)$ such that every open subset of Ω has a positive probability and the expected value of the loss

$$I(M_S) = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE \tag{2}$$

becomes a minimum for $M_S = M_S'$. (ω_E denotes that element of S for which M_{ω_E} contains E). In fact if there existed an absolutely better system M_S'' of regions, then $I(M_S'')$ would be less than $I(M_S')$ in contradiction to our assumption that $I(M_S)$ becomes a minimum for $M_S = M_S'$.

In order to obtain an admissible system M_S we may choose any bounded, uniformly continuous and everywhere positive (except if θ is contained in ω) weight function $W(\theta, \omega)$ and any arbitrary a priori distribution $f(\theta)$ (subject only to the condition that every open subset of Ω should have a positive probability) and then the system M_S which makes

$$I(M_S) = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE$$

a minimum is an admissible one. In order to determine M_S we have only to determine for each E the corresponding element ω_E of S. Let us consider the integral

$$I_E = \int_\Omega W(\theta, \omega) p(E \mid \theta)\, df(\theta).$$

The integral I_E is for a fixed E only a function of ω. It is obvious that ω_E must be that element of S for which I_E becomes a minimum.

5. **Admissible systems M_S and the Neyman-Pearson best critical regions.** Let us consider the case that the system H_S of hypotheses consists only of the following two hypotheses: 1) $\theta = \theta_0$ where θ_0 is a certain point of Ω. 2) θ belongs to the set complementary to θ_0. Let us denote by ω_1 the set consisting only of the point θ_0, and by ω_2 the set complementary to ω_1. S consists in this case only of two elements ω_1 and ω_2. The system M_S of regions consists of two regions of acceptance M_{ω_1} and M_{ω_2} corresponding to the hypotheses H_{ω_1} and H_{ω_2}. If a common best critical region in the sense of Neyman-Pearson exists and if M_S is admissible, then M_{ω_2} is obviously a common best critical region. This leads to the following remarkable conclusion: If a common best critical region exists and if the system M_S of regions consisting of the two regions M_{ω_1} and M_{ω_2} minimizes the expectation of the loss (formula 2) for a weight function and for an a priori distribution subject to some weak conditions mentioned in paragraph 4, then M_{ω_2} is a common best critical region. That is to say, the form of the weight function and of the a priori distribution affects only the size of the region M_{ω_2} but it will always be a common best critical region.

6. **The choice of M_S if a weight function is given.** We shall now consider the case in which a weight function $W(\theta, \omega)$ is given and we shall deal with the question as to how M_S in this case is to be chosen.

If the parameter point is an unknown constant and if θ denotes the true parameter point, then the expected value of the loss is given by

$$r(\theta) = \int_M W(\theta, \omega_E) p(E \mid \theta)\, dE \tag{3}$$

where the integration is to be taken over the whole sample space M and H_{ω_E} denotes the hypothesis accepted if E is the observed sample point. That is to say ω_E is that element of S for which E is contained in the region of acceptance M_{ω_E}. We shall call the expression (3) the risk of accepting a false hypothesis if θ is the true parameter point. Since we do not know the true parameter point θ, we shall have to study the risk $r(\theta)$ as a function of θ. We shall call this function the risk function. The form of the risk function depends on the system M_S of regions and on the form of the weight function. In order to express this fact, we shall denote the risk function corresponding to the system M_S and to the weight function $W(\theta, \omega)$ also by

$$r[\theta \mid M_S, W(\theta, \omega)].$$

Definition 4. Denote by M_S and M'_S two systems of regions of acceptance corresponding to the same system H_S of hypotheses. We shall say that M_S and M'_S are equivalent relative to the weight function $W(\theta, \omega)$ if the risk function

$r[\theta \mid M_S, W(\theta, \omega)]$ is identically equal to the risk function $r[\theta \mid M'_S, W(\theta, \omega)]$, that is to say if for each point θ,

$$r[\theta \mid M'_S, W(\theta, \omega)] = r[\theta \mid M_S, W(\theta, \omega)].$$

Definition 5. Denote by M_S and M'_S two systems of regions corresponding to the same system H_S of hypotheses. We shall say that M_S is uniformly better than M'_S relative to the weight function $W(\theta, \omega)$ if M_S and M'_S are not equivalent and for each θ

$$r(\theta \mid M_S, W(\theta, \omega)] \leq r[\theta \mid M'_S, W(\theta, \omega)].$$

Definition 6. A system M_S of regions of acceptance is said to be admissible relative to the weight function $W(\theta, \omega)$ if no uniformly better system of regions exists relative to the weight function considered.

It is obvious that *we have to choose a system M_S of regions which is admissible relative to the weight function considered.*

There exist in general many systems M_S which are admissible relative to the weight function given The question arises as to how can we distinguish among them. Denote by r_{M_S} the maximum of the risk function corresponding to the system M_S of regions and to the given weight function. If we do not take into consideration a priori probabilities of θ, then it seems reasonable to choose that system M_S for which r_{M_S} becomes a minimum. We shall see in section 8 that the system M_S for which r_{M_S} becomes a minimum has some important properties which justify the distinction of this particular system of regions among all admissible systems.

Definition 7. We shall call an admissible system M'_S of regions for which r_{M_S} becomes a minimum a best system of regions of acceptance relative to the weight function given.[4]

Now we shall have to deal with the question of determining a best system M_S of regions and what special properties this system M_S has.

7. Reduction of the problem to the case when the system H_S of hypotheses is the system of all simple hypotheses. A hypothesis H_ω is said to be a simple hypothesis if ω contains exactly one point of the parameter space Ω. We assume that each element ω of S is a closed subset of Ω. Hence the power of S is not greater than the power of the continuum and therefore we can always set up a correspondence between the elements ω of S and the points θ of Ω such that to each point θ corresponds a certain element ω_θ of S and to each element ω of S at least one point θ exists for which $\omega_\theta = \omega$. For instance if S consists of the two elements ω_1 and ω_2 then we can set up a correspondence as follows: the element ω_θ of S corresponding to θ is ω_1 if θ is contained in ω_1 and ω_2 otherwise.

[4] As we shall see later (Theorem 3), the best system of regions is uniquely determined if some regularity conditions are fulfilled.

If Ω is one dimensional and S is the system of all intervals of a certain length ϵ then we can define the interval ω_θ corresponding to θ as the interval of which the initial point is θ and the terminal point $\theta + \epsilon$.

Let us denote the weight function by $W(\theta, \omega)$ defined for all values of θ and for all elements ω of S. Consider the system $H_{\bar{S}}$ of all simple hypotheses and the following weight function

$$W(\theta, \bar{\theta}) = W(\theta, \omega_{\bar{\theta}}) \tag{4}$$

where θ denotes the true parameter point and $\bar{\theta}$ denotes the estimated point. A system $M_{\bar{S}}$ of regions of acceptance for $H_{\bar{S}}$ is given by a vector function $\bar{\theta}(E)$ of the observations such that to each point $E = (x_1, \cdot \cdot, x_n)$ of the sample space M corresponds a certain point $\bar{\theta}(E)$ of the parameter space. For each point θ_0 the region M_{θ_0} of the acceptance of the hypothesis $\theta = \theta_0$ is given by the equation $\bar{\theta}(E) = \theta_0$. We shall call the function $\bar{\theta}(E)$ an estimate of θ, the system of regions $M_{\bar{S}}$ is uniquely determined by the estimate. We shall call $\bar{\theta}(E)$ a best estimate relative to a given weight function if the system of regions determined by $\bar{\theta}(E)$ is a best system of regions relative to the weight function considered.

Let us denote by $\tilde{\theta}(E)$ a best estimate of θ relative to the weight function $W(\theta, \bar{\theta})$ defined in (4). A best system M_S of regions of acceptance in the original problem can obviously be obtained in the following way: Denote by ω an element of S. The region M_ω of acceptance of the hypothesis H_ω consists of the points E for which

$$\omega_{\tilde{\theta}(E)} = \omega.$$

Hence we can restrict our considerations to the case when the system of hypotheses is the system of all simple hypotheses. We shall deal with the problem of how a best estimate of θ can be found and what properties this estimate has.

8 **Some theorems concerning the best estimate.** In order to study the properties of a best estimate $\tilde{\theta}(E)$ it is useful to consider hypothetical a priori distributions of θ. We shall especially consider point distributions of θ, that is to say, distributions where a finite number of points $\theta_1, \cdot \cdot, \theta_s$ of the parameter space Ω exist such that the probability of any subset of Ω not containing any of the points $\theta_1, \cdot \cdot, \theta_s$ is zero. If $\theta_1, \cdot \cdot \cdot, \theta_s$ are given, a point distribution is characterized by a vector $\rho = (\rho_1, \cdot \cdot \cdot, \rho_s)$ where ρ_i denotes the probability of θ_i and $\Sigma\rho_i = 1$.

If $\theta(E)$ denotes an estimate of θ and if $f(\theta)$ denotes a distribution function of θ then the expected value of the loss, that is to say the expected value of the weight function $W[\theta, \theta(E)]$ is obviously given by

$$\int_M \int_\Omega W[\theta, \theta(E)]p(E \mid \theta)\, df(\theta)\, dE \tag{5}$$

where $p(E \mid \theta)$ denotes the probability density in E if θ is the true parameter point and the integration is to be taken over the product of the sample space M and parameter space Ω.

Let us assume that for every sample point E there exists a parameter point $\theta_f(E)$ such that the expression

$$\int_\Omega W(\theta, \bar{\theta}) p(E \mid \theta)\, df(\theta) \tag{6}$$

becomes a minimum with respect to $\bar{\theta}$ for $\bar{\theta} = \theta_f(E)$. We shall call the estimate $\theta_f(E)$ a minimum risk estimate with respect to the distribution $f(\theta)$, since also the expression (5) becomes a minimum for the estimate $\theta_f(E)$.

We shall make the following assumptions:

Assumption 1. The parameter space is a bounded and closed subset of the k-dimensional Euclidean space.

Assumption 2. The weight function $W(\theta, \bar{\theta})$ is continuous in θ and $\bar{\theta}$ jointly.

Assumption 3. The probability density $p(E \mid \theta)$ is continuous in E and θ jointly. That is to say if $\lim E_i = E$ and $\lim \theta_i = \theta$ then $\lim p(E_i \mid \theta_i) = p(E \mid \theta)$.

Assumption 4. For any distribution $f(\theta)$ of θ there exists at most one minimum risk estimate $\theta_f(E)$.[5]

Assumption 5. If $f(\theta)$ and $f'(\theta)$ denote two different point distributions of θ and if $\theta_f(E)$ and $\theta_{f'}(E)$ are minimum risk estimates corresponding to $f(\theta)$ and $f'(\theta)$ respectively, then $\theta_f(E)$ is not identically equal to $\theta_{f'}(E)$.

The assumptions 1–5, with addition of an assumption 6 which we shall formulate later, enables us to deduce important properties of the best estimate $\bar{\theta}(E)$. First we shall prove some Lemmas by means of the assumptions 1–5.

LEMMA 1. *For any a priori distribution $f(\theta)$ of θ there exists exactly one minimum risk estimate $\theta_f(E)$.*

According to Assumption 2 $W(\theta, \bar{\theta})$ is continuous. Since the parameter space Ω is compact on account of Assumption 1, $W(\theta, \bar{\theta})$ is uniformly continuous. According to Assumption 3 $p(E \mid \theta)$ is continuous; hence for any fixed sample point E, $p(E \mid \theta)$ is bounded. From these facts it follows easily that the expression (6) is a continuous function of $\bar{\theta}$ for any fixed sample point E. Hence there exists at least one parameter point $\theta_f(E)$ such that (6) becomes a minimum for $\bar{\theta} = \theta_f(E)$. Since, according to Assumption 4, at most one parameter point exists for which (6) becomes a minimum, Lemma 1 is proved.

If a distribution $f(\theta)$ of θ is given then the distribution of each of the components $\theta^{(1)}, \quad , \theta^{(k)}$ of θ can be found. Denote by Q_j the set of real numbers which are discontinuities of the distribution of the component $\theta^{(j)} (j = 1, \cdots, k)$ and form the set $Q = Q_1 + \cdots + Q_k$. As is well known, Q is at most denumerable. A k-dimensional interval J of the parameter space given by

$$a_j \leq \theta^{(j)} \leq b_j \qquad (j = 1, \cdots, k)$$

is called a continuity interval of the distribution $f(\theta)$ if no a_j and no b_j belongs to Q. A sequence $\{f_n(\theta)\}$ of distributions is said to be convergent towards the

[5] As will be shown in Section 10, Assumption 4 is not as restrictive as it would appear. It will be satisfied in the great majority of practical cases.

distribution $f(\theta)$, i.e. in symbols $\lim f_n(\theta) = f(\theta)$, if for any continuity interval J of $f(\theta)$ the probability of J corresponding to the distribution $f_n(\theta)$ converges with increasing n towards the probability of J corresponding to the distribution $f(\theta)$.

LEMMA 2. *If* $\{f_n(\theta)\}$ $(n = 1, \cdots, \text{ad inf.})$ *denotes a sequence of distributions, then there exists a subsequence* $\{f_{n_m}(\theta)\}$ $(m = 1, \cdots, \text{ad inf.})$ *which converges towards a distribution.*

As is well known, there exists a completely additive set function $P(\omega)$ defined for all Borel measurable subsets ω of Ω and a subsequence $\{n_m\}$ of $\{n\}$, such that for any continuity interval J of $P(\omega)$ the probability of J corresponding to the distribution $f_{n_m}(\theta)$ converges with increasing m towards $P(J)$. Since Ω is bounded, there exists a continuity interval J such that for all n the probability of J according to $f_n(\theta)$ is equal to 1. Hence $P(\Omega) = 1$, that is to say, $P(\omega)$ is a probability set function which proves Lemma 2.

LEMMA 3. *If* $\{f_n(\theta)\}$ $(n = 1, \cdots, \text{ad inf.})$ *denotes a sequence of distributions which converges towards the distribution* $f(\theta)$ *and if* $\lim E_n = E$ *then*

$$\lim_{n=\infty} \theta_{f_n}(E_n) = \theta_f(E),$$

where $\theta_{f_n}(E)$ *denotes the minimum risk estimate corresponding to* $f_n(\theta)$ *and* $\theta_f(E)$ *denotes the minimum risk estimate corresponding to* $f(\theta)$.

If $\{\varphi_n(\theta)\}$ denotes a sequence of real valued functions which converges uniformly towards a continuous function $\varphi(\theta)$ then

$$\lim \int_\Omega \varphi_n(\theta)\, df_n(\theta) = \int_\Omega \varphi(\theta)\, df(\theta). \tag{7}$$

Since $\{\varphi_n(\theta)\}$ converges uniformly towards $\varphi(\theta)$, (7) is obviously true if

$$\lim \int_\Omega \varphi(\theta)\, df_n(\theta) = \int_\Omega \varphi(\theta)\, df(\theta)$$

holds. The latter equality follows easily from the fact that Ω is compact.

Consider a subsequence $\{n_m\}$ of $\{n\}$ such that $\lim_{m=\infty} \theta_{f_{n(m)}}(E_{n_m})$ exists. Denote this limit by θ^*. In order to prove Lemma 3, we have only to show that $\theta^* = \theta_f(E)$. If $\theta_f(E) \neq \theta^*$ then on account of Assumption 4

$$\int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df(\theta) < \int_\Omega W(\theta, \theta^*)p(E \mid \theta)\, df. \tag{8}$$

$W(\theta, \bar{\theta})$ is uniformly continuous since Ω is compact. On account of Assumption 3 also $p(E \mid \theta)$ is uniformly continuous in the product of Ω with a bounded subset of the sample space. Hence

$$W[\theta, \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)$$

converges uniformly in θ towards

$$W(\theta, \theta^*)p(E \mid \theta)$$

and we have on account of (7) and (8)

$$\lim_{m=\infty} \int_\Omega W[\theta,\ \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} = \int_\Omega W(\theta,\ \theta^*)p(E \mid \theta)\, df$$

(9)

$$> \int_\Omega W[\theta,\ \theta_f(E)]p(E \mid \theta)\, df,$$

and

$$\lim_{m=\infty} \int_\Omega W[\theta,\ \theta_f(E)]p(E \mid \theta)\, df_{n_m} = \int_\Omega W[\theta,\ \theta_f(E)]p(E \mid \theta)\, df. \tag{10}$$

From (9) and (10) it follows that there exists a positive δ such that for sufficiently large m

$$\int_\Omega W[\theta,\ \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} > \int_\Omega W[\theta,\ \theta_f(E)]p(E \mid \theta)\, df_{n_m} + \delta.$$

Since the sequence of functions $\{p(E_n \mid \theta)\}$ converges uniformly in θ towards $p(E \mid \theta)$, we have for sufficiently large m

$$\int_\Omega W[\theta,\ \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} > \int_\Omega W[\theta,\ \theta_f(E)]p(E_{n_m} \mid \theta)\, df_{n_m}$$

But this is a contradiction, since $\theta_{f_n}(E)$ is a minimum risk estimate. Hence the assumption $\theta^* \neq \theta_f(E)$ is proved to be an absurdity. This proves Lemma 3.

LEMMA 4. *To each positive ϵ a bounded and closed subset M_ϵ of the sample space M can be given such that*

$$\int_{M_\epsilon} p(E \mid \theta)\, dE \geq 1 - \epsilon$$

for every point θ of the parameter space Ω.

Let us assume that Lemma 4 is not true and we shall deduce a contradiction. Denote by $M_\nu (\nu = 1, 2, \cdots, \text{ad inf.})$ the sphere in the sample space M whose center is the origin and whose radius is equal to ν. Since Lemma 4 is supposed to be not true, to each ν there exists a parameter point θ_ν such that

$$\int_{M_\nu} p(E \mid \theta_\nu)\, dE < 1 - \epsilon \qquad (\nu = 1, \cdots, \text{ad inf.}). \tag{11}$$

Since Ω is compact, there exists a subsequence $\{\theta_{\nu_\mu}\}$ of the sequence $\{\theta_\nu\}$ such that $\lim_{\mu=\infty} \theta_{\nu_\mu}$ exists. Denote $\lim_{\mu=\infty} \theta_{\nu_\mu}$ by θ. Since

$$\int_M p(E \mid \theta)\, dE = 1$$

there exists a positive integer ν' such that

$$\int_{M_{\nu'}} p(E \mid \theta)\, dE > 1 - \frac{\epsilon}{2}.$$

On account of Assumption 3 we get easily

$$\lim_{\mu=\infty} \int_{M_{\nu'}} p(E \mid \theta_{\nu_\mu}) \, dE = \int_{M_{\nu'}} p(E \mid \theta) \, dE.$$

Hence for sufficiently large μ we get

$$\int_{M_{\nu_\mu}} p(E \mid \theta_{\nu_\mu}) \, dE \geq \int_{M_{\nu'}} p(E \mid \theta_{\nu_\mu}) \, dE > 1 - \epsilon,$$

in contradiction to (11). This proves Lemma 4.

For any estimate $\theta(E)$ we shall call the integral

$$r(\theta) = \int_M W[\theta, \theta(E)] p(E \mid \theta) \, dE$$

the risk function of the estimate $\theta(E)$. The value of the risk function $r(\theta)$ is for any θ equal to the expected value of the loss (of the weight function) if θ is the true parameter point.

LEMMA 5. *To any positive η a positive δ can be given such that for any estimate $\theta(E)$ and for any pair θ, θ' of parameter points whose Euclidean distance is less than δ the inequality*

$$| r(\theta) - r(\theta') | = \left| \int_M W[\theta, \theta(E)] p(E \mid \theta) \, dE - \int_M W[\theta', \theta(E)] p(E \mid \theta') \, dE \right| < \eta$$

holds.

Since $W(\theta, \bar{\theta})$ is uniformly continuous, to any $\epsilon > 0$ a positive δ can be given such that for any pair of points θ, θ' whose Euclidean distance is less than δ the relation

$$| W(\theta, \bar{\theta}) - W(\theta', \bar{\theta}) | < \epsilon \tag{12}$$

holds for every $\bar{\theta}$. On account of Assumption 3 δ can be chosen in such a way that also the inequality

$$| p(E \mid \theta) - p(E \mid \theta') | < \epsilon \tag{13}$$

is satisfied for any sample point E of a bounded subset M' of M and for any pair θ, θ' whose Euclidean distance is less than δ.

Since $W(\theta, \bar{\theta})$ is continuous and Ω is compact, $W(\theta, \bar{\theta})$ must be bounded. Denote by A an upper bound of $W(\theta, \bar{\theta})$. According to Lemma 4 there exists a bounded and closed subset M' of the sample space M such that

$$\int_{M'} p(E \mid \theta) \, dE \geq 1 - \frac{\eta}{2A} \text{ for any } \theta.$$

It is obvious that

$$\left| \int_{M-M'} W[\theta, \theta(E)] p(E \mid \theta) \, dE - \int_{M-M'} W[\theta', \theta(E)] p(E \mid \theta') \, dE \right| \leq \frac{\eta}{2}.$$

In order to prove Lemma 5 we have only to show that

$$\left| \int_{M'} W[\theta, \theta(E)]p(E \mid \theta)\, dE - \int_{M'} W[\theta', \theta(E)]p(E \mid \theta')\, dE \right| < \frac{\eta}{2}. \tag{14}$$

On account of (12) and (13), (14) is certainly true for sufficiently small ϵ. Hence Lemma 5 is proved.

LEMMA 6. *If the sequence $\{f_n(\theta)\}$ of distributions converges towards the distribution $f(\theta)$ and if $r_{f_n}(\theta)$ denotes the risk function of the minimum risk estimate $\theta_{f_n}(E)$ then $\{r_{f_n}(\theta)\}$ converges uniformly towards the risk function $r_f(\theta)$ of the minimum risk estimate $\theta_f(E)$.*

According to Lemma 4 to any positive ϵ a bounded and closed subset M_ϵ of M can be given such that

$$\int_{M_\epsilon} p(E \mid \theta)\, dE \geq 1 - \epsilon \tag{15}$$

for every θ. From Lemma 3 it follows easily that $\{\theta_{f_n}(E)\}$ converges uniformly towards $\theta_f(E)$ in M_ϵ. Hence

$$\lim_{n=\infty} \int_{M_\epsilon} W[\theta, \theta_{f_n}(E)]p(E \mid \theta)\, dE = \int_{M_\epsilon} W[\theta, \theta_f(E)]p(E \mid \theta)\, dE$$

holds for every θ and for every positive ϵ. Since $W(\theta, \bar{\theta})$ is bounded and ϵ can be chosen arbitrarily small, we get on account of (15) that

$$\lim_{n=\infty} \int_{M} W[\theta, \theta_{f_n}(E)]p(E \mid \theta)\, dE = \int_{M} W[\theta, \theta_f(E)]p(E \mid \theta)\, dE,$$

that is to say

$$\lim r_{f_n}(\theta) = r_f(\theta).$$

The uniformity of the convergence follows easily from Lemma 5.

In the following argument we shall consider an arbitrary but fixed system of s parameter points $\theta_1, \cdots, \theta_s$, and point distributions such that no point $\theta \neq \theta_1, \cdots, \theta_s$ has positive probability. Such a point distribution is characterized by a vector $\rho = (\rho_1, \cdots, \rho_s)$ where ρ_i denotes the probability of θ_i $(i = 1, \cdots, s)$ and $\Sigma\rho_i = 1$. The points $\theta_1, \cdots, \theta_s$ are kept constant and only ρ will vary. Hence if we speak about different distributions $\rho = (\rho_1, \cdots, \rho_s)$, $\rho' = (\rho_1', \cdots, \rho_s')$ they are always related to the same points $\theta_1, \cdots, \theta_s$ unless we state explicitly the contrary.

LEMMA 7. *If $\rho = (\rho_1, \cdots, \rho_s)$ and $\rho' = (\rho_1 + \Delta\rho_1, \cdots, \rho_s + \Delta\rho_s)$ denote two different distributions then*

$$\sum_{i=1}^{s} [(\lambda - 1)\rho_i + \lambda\Delta\rho_i][r_i(\rho') - r_i(\rho)] < 0$$

holds for any positive λ, *where*

$$r_i(\rho) = \int_M W[\theta_i, \theta_\rho(E)]p(E \mid \theta_i)\, dE \qquad (i = 1, \cdots, s),$$

$$r_i(\rho') = \int_M W(\theta_i, \theta_{\rho'}(E)]p(E \mid \theta_i)\, dE,$$

and $\theta_\rho(E)$ *and* $\theta_{\rho'}(E)$ *denote the minimum risk estimates corresponding to* ρ *and* ρ' *respectively.*

We have

$$\sum_i (\rho_i + \Delta\rho_i)r_i(\rho) = \int_M \sum_i W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i)\, dE = I_1$$

and

$$\sum_i (\rho_i + \Delta\rho_i)r_i(\rho') = \int_M \sum W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i)\, dE = I_2.$$

Since $\theta_{\rho'}(E)$ is the minimum risk estimate corresponding to ρ', we have $I_1 \geq I_2$. We shall show that $I_1 > I_2$. According to Assumption 5 $\theta_\rho(E)$ is not identically equal to $\theta_{\rho'}(E)$. Hence there exists a point E' such that $\theta_\rho(E') \neq \theta_{\rho'}(E')$. On account of Assumption 4

$$\Sigma W[\theta_i, \theta_\rho(E')]\rho_i' p(E' \mid \theta_i) > \Sigma W[\theta_i, \theta_{\rho'}(E')]\rho_i' p(E' \mid \theta_i).$$

From Lemma 3 it follows that $\theta_\rho(E)$ and $\theta_{\rho'}(E)$ are continuous functions of E. Hence there exists a positive δ and a sphere s with center in E' such that

$$\Sigma W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i) > \Sigma W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i) + \delta$$

for every point E of S. Since $\theta_{\rho'}(E)$ is the minimum risk estimate corresponding to ρ' we have

$$\Sigma W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i) \geq \Sigma W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i)$$

for every point E outside S. Hence $I_1 > I_2$ that is to say

$$\Sigma(\rho_i + \Delta\rho_i)r_i(\rho) > \Sigma(\rho_i + \Delta\rho_i)r_i(\rho'). \tag{16}$$

Analogously we get

$$\Sigma\rho_i r_i(\rho) < \Sigma\rho_i r_i(\rho'). \tag{17}$$

Multiplying (16) by an arbitrary positive value λ and subtracting (17) we get

$$\Sigma[\lambda(\rho_i + \Delta\rho_i) - \rho_i]r_i(\rho) > \Sigma[\lambda(\rho_i + \Delta\rho_i) - \rho_i]r_i(\rho').$$

Hence

$$\Sigma[(\lambda - 1)\rho_i + \lambda\Delta\rho_i][r_i(\rho') - r_i(\rho)] < 0.$$

Let us denote for any ρ the maximum of the numbers

by $r(\rho)$. We shall call a distribution ρ for which $r(\rho)$ becomes a minimum, a risk-minimizing distribution. We shall say that the risk-minimizing distribution $\rho = (\rho_1, \cdots, \rho_s)$ is not degenerate if $\rho_1 > 0, \cdots, \rho_s > 0$. Otherwise we shall say that ρ is degenerate.

LEMMA 8. *There exists at least one risk-minimizing distribution* ρ.

From Lemma 6 it follows that $r_1(\rho), \cdots, r_s(\rho)$ are continuous functions of ρ. Hence also $r(\rho)$ is continuous Since the set of all possible distributions ρ is bounded and closed, there must be at least one distribution ρ for which $r(\rho)$ becomes a minimum.

LEMMA 9. *If* $\rho = (\rho_1, \cdots, \rho_s)$ *denotes a risk-minimizing distribution which is not degenerate then*

$$r_1(\rho) = r_2(\rho) = \cdots = r_s(\rho).$$

Let us assume that there are two integers i and j, for instance 1 and 2, such that $r_1(\rho) < r_2(\rho)$. We shall deduce a contradiction from this assumption. Let us consider two different distributions $\rho' = (\rho_1', \cdots, \rho_s')$ and $\rho'' = (\rho_1'', \cdots, \rho_s'')$ where $\rho_1'' > 0$. Hence at least one of the quantities

$$(\rho_1' - \rho_1''), \cdots, (\rho_s' - \rho_s'')$$

is unequal to zero. Since $\Sigma\rho_i' = \Sigma\rho_i'' = 1$, also at least one of the quantities

$$(\rho_2' - \rho_2''), \cdots, (\rho_s' - \rho_s'')$$

must be unequal to zero. On account of Lemma 7 we have

$$\sum_{i=1}^{s} [(\lambda - 1)\rho_i' + \lambda(\rho_i'' - \rho_i')][r_i(\rho'') - r_i(\rho')] < 0.$$

If we put $\lambda = \frac{\rho_1'}{\rho_1''}$ we get

$$\sum_{i=2}^{s} \left[\left(\frac{\rho_1'}{\rho_1''} - 1\right)\rho_i' + \frac{\rho_1'}{\rho_1''}(\rho_i'' - \rho_i')\right][r_i(\rho'') - r_i(\rho')] < 0.$$

Hence at least one of the quantities

$$r_2(\rho'') - r_2(\rho'), \cdots, r_s(\rho'') - r_s(\rho')$$

must be unequal to zero.

Since $\rho_1 > 0$, there exists a closed sphere S_ρ with center at ρ such that for any point ρ' of S_ρ $\rho_1' > 0$. Hence for any two different points ρ' and ρ'' of S_ρ at least one of the quantities

$$r_2(\rho'') - r_2(\rho'), \cdots, r_s(\rho'') - r_s(\rho')$$

is unequal to zero. Denote by $\bar{S}_\rho$ the projection of S_ρ on the $s - 1$ dimensional space given by $\rho_1 = 0$. Consider the transformation according to which the image of the point $\bar{\rho}' = (\rho_2', \cdots, \rho_s')$ of $\bar{S}_\rho$ is the point $\bar{q}(\bar{\rho}') = [r_2(\rho'), \cdots, r_s(\rho')]$. It is obvious that the images of two different points of $\bar{S}_\rho$ are different.

Since $r_i(\rho)$ $(i = 1, \cdots, s)$ is continuous, the transformation is continuous and therefore topological. Denote the image of $\bar{S}_\rho$ by $\bar{R}_\rho$. Since $\bar{\rho} = (\rho_2, \cdots, \rho_s)$ is an interior point of $\bar{S}_\rho$, according to the Brouwer-Jordan theorem[6] on domain invariance the image $\bar{q}(\bar{\rho}) = [r_2(\rho), \cdots, r_s(\rho)]$ of $\bar{\rho}$ must also be an interior point of $\bar{R}_\rho$. Hence for sufficiently small $\epsilon > 0$ the point

$$t(\epsilon) = [r_2(\rho) - \epsilon, \cdots, r_s(\rho) - \epsilon]$$

is contained in $\bar{R}_\rho$. Denote by $\bar{\rho}(\epsilon) = [\rho_2(\epsilon), \cdots, \rho_s(\epsilon)]$ the point of $\bar{S}_\rho$ whose image is $t(\epsilon)$. It is obvious that

$$\lim_{\epsilon=0} \bar{\rho}(\epsilon) = \bar{\rho} = (\rho_2, \cdots, \rho_s). \tag{18}$$

Consider the point $\rho(\epsilon)$ of S_ρ whose projection is $\bar{\rho}(\epsilon)$ that is to say $\rho(\epsilon)$ has the co-ordinates $1 - \Sigma\bar{\rho}_i(\epsilon)$, $\bar{\rho}_2(\epsilon), \cdots, \bar{\rho}_s(\epsilon)$. From (18) it follows that also

$$\lim_{\epsilon=0} \rho(\epsilon) = \rho = (\rho_1, \rho_2, \cdots, \rho_s). \tag{19}$$

Since $r_1[\rho(\epsilon)], \cdots, r_s[\rho(\epsilon)]$ are continuous functions of ϵ and since $r_1(\rho) < r_2(\rho)$, for sufficiently small ϵ the maximum of the numbers

$$r_1[\rho(\epsilon)], r_2[\rho(\epsilon)] = r_2(\rho) - \epsilon, \cdots, r_s[\rho(\epsilon)] = r_s(\rho) - \epsilon$$

is certainly smaller than the maximum $r(\rho)$ of the numbers

$$r_1(\rho), \cdots, r_s(\rho),$$

in contradiction to our assumption that ρ is a risk minimizing distribution. Hence the assumption $r_1(\rho) < r_2(\rho)$ is proved to be an absurdity and Lemma 9 is proved.

In the previous arguments we have considered an arbitrary but fixed system of s parameter points $\theta_1, \cdots, \theta_s$ and all distributions ρ were related to these points. In the following arguments we shall vary the points $\theta_1, \cdots, \theta_s$ and therefore we shall have to state the parameter points to which the distribution ρ is related.

Let us consider a sequence $\{\theta_\nu\}$ $(\nu = 1, \cdots, \text{ad inf.})$ of parameter points which is dense in Ω. We say that a subset ω of Ω is dense in Ω if for each point θ of Ω any arbitrarily small open neighborhood of θ contains at least one point of ω Since Ω is compact, a sequence $\{\theta_\nu\}$ which is dense in Ω certainly exists. Let us consider the first s points $\theta_1, \cdots, \theta_s$ of the sequence $\{\theta_\nu\}$. According to Lemma 8 there exists for any s a risk-minimizing distribution $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ related to $\theta_1, \cdots, \theta_s$.

Assumption 6. There exists a sequence $\{\theta_s\}$ $(s = 1, \cdots, \text{ad inf.})$ of parameter points which is dense in Ω and such that for almost any s[7] the risk-minimizing

[6] See for instance Alexandroff and Hopf, *Topologie*, Berlin 1935, p. 396

[7] By "almost any s" we understand "for all s greater than a sufficiently large integer."

distribution $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ related to the first s points $\theta_1, \cdots, \theta_s$, is not degenerate.

LEMMA 10. *Denote by* $\{\theta_s\}$ $(s = 1, 2, \cdots, \text{ad inf.})$ *a sequence of parameter points for which the conditions of Assumption* 6 *are fulfilled. Denote by* $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ *the risk-minimizing distribution related to the first* s *points* $\theta_1, \cdots, \theta_s$. *Then there exists a non-negative constant* c *such that for any arbitrarily small positive* ϵ *the inequality*

$$c - \epsilon \leq \int_M W[\theta, \theta_{\rho(s)}(E)]p(E \mid \theta)\, dE \leq c + \epsilon$$

holds identically in θ *for almost every* s. *That is to say the risk function of the minimum risk estimate* $\theta_{\rho(s)}(E)$ *lies entirely between* $c - \epsilon$ *and* $c + \epsilon$ *for almost every* s.

Denote the risk function

$$\int_M W[\theta, \theta_{\rho(s)}(E)]p(E \mid \theta)\, dE$$

of the estimate $\theta_{\rho(s)}(E)$ by $r(\theta, s)$. First we shall prove that there exists a sequence $\{c_s\}$ $(s = 1, \cdots, \text{ad inf.})$ of non-negative numbers such that for every $\epsilon > 0$ the inequality

$$c_s - \epsilon \leq r(\theta, s) \leq c_s + \epsilon \tag{20}$$

holds for almost every s. In fact to any positive η a positive integer s_η can be given such that for any $s > s_\eta$ the points $\theta_1, \cdots, \theta_s$ are η-dense in Ω. That is to say every point θ of Ω lies in a sphere with radius η and center in one of the points $\theta_1, \cdots, \theta_s$. Since for sufficiently large s $\rho(s)$ is not degenerate, we have on account of Lemma 9 for sufficiently large s

$$r(\theta_1, s) = \cdots = r(\theta_s, s) = c_s. \tag{21}$$

Since for sufficiently large s $\theta_1, \cdots, \theta_s$ is η-dense in Ω, we get easily from Lemma 5 that (20) holds for any positive ϵ for almost every s.

In order to prove Lemma 10 we have only to show that $\lim_{s=\infty} c_s$ exists and is finite. First we see that for no estimate $\theta(E)$ can the corresponding risk function

$$r(\theta) = \int_M W[\theta, \theta(E)]p(E \mid \theta)\, dE$$

lie entirely below $r(\theta, s)$ that is to say

$$r(\theta) < r(\theta, s) \tag{22}$$

cannot hold for any θ. In fact if (22) were true for a certain estimate $\theta(E)$ then

$$\Sigma\rho_i(s)r(\theta_i) = \int_M \Sigma W[\theta_i, \theta(E)]\rho_i(s)p(E \mid \theta_i)\, dE < \Sigma\rho_i(s)r(\theta_i, s)$$

$$= \int_M \Sigma W[\theta_i, \theta_{\rho(s)}(E)]\rho_i(s)p(E \mid \theta_i)\, dE,$$

which is not possible since $\theta_{\rho(\varepsilon)}(E)$ is a minimum risk estimate. Hence (22) cannot hold for any θ. From this fact follows easily that $\lim c_\varepsilon$ exists and is finite. This proves Lemma 10.

LEMMA 11. *Denote $f(\theta)$ a distribution of θ and let $\theta_f(E)$ be the corresponding minimum risk estimate. If $\theta(E)$ denotes an arbitrary estimate then*

$$r(\theta) \equiv r_f(\theta)$$

if $\theta_f(E) \neq \theta(E)$ only in a set of measure 0, and

$$\int_\Omega r(\theta)\, df(\theta) > \int_\Omega r_f(\theta)\, df(\theta)$$

if $\theta_f(E) \not\equiv \theta(E)$ in a set of positive measure. $r(\theta)$ denotes the risk function of $\theta(E)$ and $r_f(\theta)$ denotes the risk function of $\theta_f(E)$.

If $\theta_f(E) \not\equiv \theta(E)$ only in a set of measure zero, then we have obviously $r(\theta) \equiv r_f(\theta)$. Consider the case that $\theta_f(E) \not\equiv \theta(E)$ in a set M' of positive measure. According to Assumption 4 we have

$$\int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df(\theta) > \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df(\theta)$$

for any point E of M'. Since

$$\int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df(\theta) = \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df(\theta)$$

for any other point E of the sample space M, we get

$$\int_\Omega r(\theta)\, df = \int_M \int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df\, dE$$

$$> \int_M \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df\, dE = \int_\Omega r_f(\theta)\, df.$$

Hence Lemma 11 is proved.

We are now able to prove some theorems about the best estimate $\bar{\theta}(E)$ relative to a given weight function. An estimate $\bar{\theta}(E)$ is a best estimate according to our definition 7, if the maximum of the risk function of $\bar{\theta}(E)$ is less than or equal to the maximum of the risk function of any other estimate $\theta(E)$ and if $\bar{\theta}(E)$ is an admissible estimate (that is to say there exists no estimate $\theta(E)$ such that the risk function $r(\theta)$ of $\theta(E)$ is not identical to the risk function $\bar{r}(\theta)$ of $\bar{\theta}(E)$ and in every point θ $\bar{r}(\theta) \geq r(\theta)$.

THEOREM 1. *If $\bar{\theta}(E)$ is a best estimate and if the Assumptions 1–6 are fulfilled then the risk function $\bar{r}(\theta)$ of $\bar{\theta}(E)$ is constant, that is to say*

$$\bar{r}(\theta) \equiv c.$$

According to Assumption 6 there exists a sequence $\{\theta_s\}$ ($s = 1, \cdots$, ad inf.) of parameter points such that $\{\theta_s\}$ is dense in Ω and for almost every s the risk-

minimizing distribution $\rho(s)$ related to $\theta_1, \cdots, \theta_s$ is not degenerate. On account of Lemma 10 there exists a non-negative constant c such that for any $\epsilon > 0$ the inequality

$$c - \epsilon \leq r(\theta, s) \leq c + \epsilon \tag{23}$$

holds for almost every s. $r(\theta, s)$ denotes the risk function of the estimate $\theta_{\rho(s)}(E)$. According to Lemma 2 there exists a subsequence $\{s_n\}$ $(n = 1, \cdots,$ ad inf.) of integers such that the sequence $\{\rho(s_n)\}$ of distributions converges towards a distribution $f(\theta)$. From Lemma 6 it follows that

$$\lim_{n=\infty} r(\theta, s_n) = r_f(\theta)$$

where $r_f(\theta)$ denotes the risk function of the minimum risk estimate $\theta_f(E)$. On account of (23) we have

$$r_f(\theta) \equiv c.$$

From Lemma 11 it follows that for any other estimate $\theta(E)$ either

$$r(\theta) \equiv r_f(\theta) \equiv c$$

or

$$\int_\Omega r(\theta)\, df > \int_\Omega r_f(\theta)\, df,$$

where $r(\theta)$ denotes the risk function of $\theta(E)$. In the latter case there exists at least one point θ for which $r(\theta) > r_f(\theta)$. Hence $\theta_f(E)$ is a best estimate. If $\bar{\theta}(E)$ is also a best estimate, we get on account of Lemma 11 that $\bar{\theta}(E)$ can differ from $\theta_f(E)$ only in a set of measure 0 and the risk function of $\bar{\theta}(E)$ is identically equal to c. Hence we have proved Theorem 1 and also the following Theorems 2–3:

THEOREM 2. *If the Assumptions* 1–6 *are fulfilled there exists a distribution* $f(\theta)$ *of* θ *such that the corresponding minimum risk estimate* $\theta_f(E)$ *is a best estimate.*

THEOREM 3. *If Assumptions* 1–6 *are fulfilled and* $\bar{\theta}(E)$, $\theta^*(E)$ *are best estimates, then* $\bar{\theta}(E) = \theta^*(E)$ *almost everywhere and the corresponding risk functions are identically equal.*

Now we shall prove (without making the Assumptions 1–6)

THEOREM 4. *If* $W(\theta, \bar{\theta})$ *and* $p(E \mid \theta)$ *are continuous and* Ω *is compact, and if* $f(\theta)$ *denotes a distribution of* θ *such that any open set has a positive probability, then the minimum risk estimate* $\theta_f(E)$ *is a best estimate if its risk function* $r_f(\theta)$ *is identically equal to a constant.*

Let $r_f(\theta)$ be identically equal to c and consider an arbitrary estimate $\theta(E)$. Since $W(\theta, \bar{\theta})$ and $p(E \mid \theta)$ are continuous and Ω is compact, the risk function $r(\theta)$ of $\theta(E)$ is a continuous function of θ. Since $\theta_f(E)$ is a minimum risk estimate we have

$$\int_\Omega r(\theta)\, df \geq \int_\Omega r_f(\theta)\, df = c. \tag{24}$$

In order to prove Theorem 4, we have to show that either

$$r(\theta) \equiv c \tag{25}$$

or there exists a point θ' such that

$$r(\theta') > c. \tag{26}$$

If (25) does not hold there exists a point θ^* such that $r(\theta^*) \neq c$. If $r(\theta^*) > c$ our statement is proved. Consider the case $r(\theta^*) < c$. On account of the continuity of $r(\theta)$ there exists a positive δ and an open neighborhood U of θ^* such that

$$r(\theta) < c - \delta$$

for every θ in U. Since $\int_U df$ is assumed to be positive, the inequality (24) can hold only if there exists at least a point θ' for which $r(\theta') > c$. This proves Theorem 4.

9. **Determination of the best estimate $\bar{\theta}(E)$ for a certain class of distributions $p(E \mid \theta)$.** In this paragraph we shall prove two theorems which enable us to calculate very easily the best estimate $\bar{\theta}(E)$ for a certain special but important class of distributions.

The risk function of an estimate $\bar{\theta}(E)$ is given by

$$r(\theta) = \int_M W[\theta, \bar{\theta}(E)]\, p(E \mid \theta)\, dE,$$

where the integration is to be taken over the whole sample space M. We consider the integral equation

$$\int_M W[\theta, \bar{\theta}(E)]\, p(E \mid \theta)\, dE \equiv c, \tag{27}$$

where c denotes an arbitrary constant. If we can find an estimate $\bar{\theta}(E)$ which satisfies (27) for a certain c and which is an admissible estimate relative to the weight function considered, then $\bar{\theta}(E)$ is certainly a best estimate. If Assumptions 1–6 are fulfilled, an admissible estimate satisfying (27) certainly exists. As we shall see, a best estimate can very easily be determined by the above procedure if the conditions in the following theorem 5 are fulfilled.

THEOREM 5. *Let us assume that the following conditions are fulfilled:*

I. *The parameter space Ω is one dimensional and θ can take any real value from $-\infty$ to $+\infty$.*

II. *The probability density $p(E \mid \theta)$ depends only on the differences $x_1 - \theta, \cdots, x_n - \theta$, that is to say $p(E \mid \theta) = p(x_1 - \theta, \cdots, x_n - \theta)$, where $x_1, \cdots, x_n$ denote the co-ordinates of E.*

III. *The value of the weight function depends only on the difference $u = \theta - \bar{\theta}$ and is uniformly continuous in u.*

IV. *For any value $\bar{\theta}$ and for any sample point E the integral*

$$\psi(\bar{\theta}, E) = \int_{-\infty}^{+\infty} W(\theta - \bar{\theta}) p(E \mid \theta)\, d\theta \tag{28}$$

has a finite value.

V. *For every E there exists a finite value $\theta'(E)$ such that $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \theta'(E)$.*

Then there exists an estimate $\bar{\theta}(E)$ such that for any E, $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \bar{\theta}(E)$ and $\bar{\theta}(E'') - \bar{\theta}(E') = \lambda$ for any $E' = (x_1', \cdots, x_n')$ and $E'' = (x_1'', \cdots, x_n'')$ for which $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. An estimate with these properties is a best estimate.

Let us consider two sample points $E' = (x_1', \cdots, x_n')$ and $E'' = (x_1'', \cdots, x_n'')$ such that $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. From the conditions II and III follows that if $\psi(\bar{\theta}, E')$ becomes a minimum for $\bar{\theta} = \theta_1$, then $\psi(\bar{\theta}, E'')$ becomes a minimum for $\bar{\theta} = \theta_2 = \theta_1 + \lambda$. Hence there exists an estimate $\bar{\theta}(E) = \bar{\theta}(x_1, \cdots, x_n)$ such that for any E, $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \bar{\theta}(E)$ and $\bar{\theta}(E'') - \bar{\theta}(E') = \lambda$ if $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. We shall show that such an estimate $\bar{\theta}(E)$ is a best estimate. First we shall show that the risk function

$$r(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta - \bar{\theta}(E)]\, p(x_1 - \theta, \cdots, x_n - \theta')\, dx_1 \cdots dx_n$$

is constant. Let us consider two arbitrary parameter values θ' and θ''. Then we have

$$r(\theta') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta' - \bar{\theta}(E)]\, p(x_1 - \theta', \cdots, x_n - \theta')\, dx_1 \cdots dx_n,$$

$$r(\theta'') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta'' - \bar{\theta}(E)]\, p(x_1 - \theta'', \cdots, x_n - \theta'')\, dx_1 \cdots dx_n.$$

Making in the second integral the transformation

$$y_1 = x_1 - (\theta'' - \theta'), \cdots, y_n = x_n - (\theta'' - \theta'),$$

we get

$$r(\theta'') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W\{\theta'' - \bar{\theta}[y_1 + (\theta'' - \theta'), \cdots, y_n + (\theta'' - \theta')]\}\, p(y_1 - \theta', \cdots, y_n - \theta')\, dy_1 \cdots dy_n$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta' - \bar{\theta}(y_1, \cdots, y_n)]\, p(y_1 - \theta', \cdots, y_n - \theta')\, dy_1 \cdots dy_n.$$

Hence $r(\theta') = r(\theta'')$ and our statement that $r(\theta)$ is constant is proved. In order to prove Theorem 5, we have only to show that $\bar{\theta}(E)$ is an admissible estimate. For this purpose let us consider an arbitrary estimate $\theta^*(E)$ and

denote the corresponding risk function by $r^*(\theta)$. Since $\bar{\theta}(E)$ minimizes the integral (28), we have

$$\psi[\theta^*(E), E] \geq \psi[\bar{\theta}(E), E] \tag{29}$$

for all sample points E. Let us consider the integral

$$I = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \{W[\theta - \bar{\theta}(E)] - W[\theta - \theta^*(E)]\} p(E \mid \theta)\, d\theta\, dx_1 \cdots dx_n. \tag{30}$$

Integrating (30) with respect to θ we get

$$I = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \{\psi[\bar{\theta}(E), E] - \psi[\theta^*(E) - E]\}\, dx_1 \cdots dx_n. \tag{31}$$

Integrating (30) with respect to E, we get

$$I = \int_{-\infty}^{+\infty} [r(\theta) - r^*(\theta)]\, d\theta. \tag{32}$$

On account of (29) and (31) we have $I \leq 0$, hence

$$\int_{-\infty}^{+\infty} [r(\theta) - r^*(\theta)]\, d\theta \leq 0. \tag{33}$$

From (33) it follows that if $r^*(\theta) \leq r(\theta)$ for every θ then $r^*(\theta) < r(\theta)$ can hold only for the points of a set of measure zero. In case $r^*(\theta)$ is continuous, this means that $r^*(\theta) \equiv r(\theta)$. Hence if $r^*(\theta)$ is continuous, then either $r^*(\theta) \equiv r(\theta)$ or there exists at least one point θ' such that $r^*(\theta') > r(\theta)$. The risk function $r^*(\theta)$ is continuous if the estimate $\theta^*(E)$ is uniformly continuous in the whole sample space. In fact, we have

$$r^*(\theta + t) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta + t - \theta^*(E)]\, p(x_1 - \theta - t, \cdots, x_n - \theta - t)\, dx_1 \cdots dx_n.$$

Making the transformation

$$y_i = x_i - t \qquad (i = 1, \cdots, n)$$

we get

$$r^*(\theta + t) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta + t - \theta^*(y_1 + t, \cdots, y_n + t)]\, p(y_1 - \theta, \cdots, y_n - \theta)\, dy_1 \cdots dy_n.$$

Since $W(u)$ and $\theta^*(E)$ are uniformly continuous, from the latter equation we get easily

$$\lim_{t=0} r^*(\theta + t) = r^*(\theta)$$

that is to say $r^*(\theta)$ is continuous. Considering only continuous estimates the admissibility of $\bar{\theta}(E)$, and therefore also Theorem 5, is proved. If $\theta^*(E)$ is not

uniformly continuous we have only proved that if $r^*(\theta) \leq r(\theta)$ for every θ, then $r^*(\theta) < r(\theta)$ can hold only in a set of measure zero. I should like to mention without proof that even if $\theta^*(E)$ is not continuous, $r^*(\theta) \leq r(\theta)$ implies $r^*(\theta) \equiv r(\theta)$.

An estimate $\hat{\theta}(E)$ is called a maximum likelihood estimate if for any fixed E $p(E \mid \theta)$ becomes a maximum with respect to θ for $\theta = \hat{\theta}(E)$.

THEOREM 6. *Consider the following conditions*:

VI. *There exists exactly one maximum likelihood estimate $\hat{\theta}(E)$ with the following properties*:

a) *For any E $p(E \mid \theta)$ is non-decreasing with increasing θ for $\theta < \hat{\theta}(E)$ and non-increasing with increasing θ for $\theta > \hat{\theta}(E)$.*

b) *For any E $p(E \mid \theta)$ is a symmetric function of θ about $\hat{\theta}(E)$ that is to say, for for any real value λ $p[E \mid \hat{\theta}(E) - \lambda] = p[E \mid \hat{\theta}(E) + \lambda]$.*

VII. *The value of the weight function depends only on the absolute value of the difference $u = \theta - \bar{\theta}$ and $\frac{dw(u)}{du}$ exists, is uniformly continuous and > 0 for $u > 0$.*

If the conditions I–V of Theorem 5 and the above condition VII are fulfilled, and if $\hat{\theta}(E)$ is a maximum likelihood estimate satisfying VI, then $\hat{\theta}(E)$ is a best estimate

Assume that the conditions I–V and VII are satisfied and that $\hat{\theta}(E)$ is a maximum likelihood estimate satisfying VI. It is obvious that $\hat{\theta}(E'') - \hat{\theta}(E') = \lambda$ for $E' = (x_1, \cdots, x_n)$ and $E'' = (x_1 + \lambda, \cdots, x_n + \lambda)$. In order to prove Theorem 6, we have, according to Theorem 5, only to show that the integral in (28)

$$\psi(\bar{\theta}, E) = \int_{-\infty}^{+\infty} W(\theta - \bar{\theta}) p(E \mid \theta)\, d\theta$$

becomes a minimum for $\bar{\theta} = \hat{\theta}(E)$. Denote $\theta - \bar{\theta}$ by u. Since $\frac{dW(u)}{du}$ is uniformly continuous, we have

$$\frac{\partial \psi(\bar{\theta}, E)}{\partial \bar{\theta}} = \int_{-\infty}^{+\infty} -\left[\frac{dW(u)}{du}\right] p(E \mid \theta)\, d\theta.$$

Since $\frac{dW(u)}{du} = -\frac{dW(-u)}{du}$ we have

$$\frac{\partial \psi(\bar{\theta}, E)}{\partial \bar{\theta}} = \int_0^{\infty} \left[\frac{dW(u)}{du}\right] [p(E \mid \bar{\theta} - u) - p(E \mid \bar{\theta} + u)]\, du. \tag{34}$$

From condition VI it follows easily that for any fixed E and $\bar{\theta}$ the function of u $(0 \leq u \leq \infty)$

$$p(E \mid \bar{\theta} - u) - p(E \mid \bar{\theta} + u)$$

does not change its sign and if $\bar{\theta} \neq \hat{\theta}(E)$ there exists an interval J such that the above expression is unequal to zero for every point u of J. Hence on account of $\frac{dW(u)}{du} > 0$ for $u > 0$, the integral in (34) vanishes only for $\bar{\theta} = \hat{\theta}(E)$. Since according to the condition V there exists a finite value θ' such that $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \theta'$, θ' must be equal to $\hat{\theta}(E)$. This proves Theorem 6.

The condition VI is seldom exactly fulfilled. But for large n, in the great majority of practical cases, VI will be fulfilled with good approximation and the best estimate approaches the maximum likelihood estimate with increasing n.

10. **Two examples.** As a first example we consider a normal distribution with the variance 1. The mean value θ is unknown and we have to estimate it by means of a sample $E = (x_1, \cdots, x_n)$. In this case

$$p(E \mid \theta) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\Sigma(x_i-\theta)^2}.$$

It is obvious that for a very broad class of weight functions the conditions I–V of Theorem 5 are fulfilled. The maximum likelihood estimate $\hat{\theta}(x_1, \cdots, x_n) = \frac{x_1 + \cdots + x_n}{n}$ satisfies the condition VI of Theorem 6. Hence if the weight function satisfies also the condition VII, then the best estimate of θ is the maximum likelihood estimate $\hat{\theta}(x_1, \cdots, x_n) = \frac{x_1 + \cdots + x_n}{n}$.

Let us now consider a weight function defined as follows:

$$W(\theta, \bar{\theta}) = 2(\bar{\theta} - \theta) \quad \text{if} \quad \bar{\theta} \geq \theta$$

and

$$W(\theta, \bar{\theta}) = \theta - \bar{\theta} \qquad \text{if} \quad \bar{\theta} < \theta.$$

Since for this weight function, the conditions I–V satisfied, according to Theorem 5 the best estimate of θ is the value $\bar{\theta}$ for which the integral

$$\int_{-\infty}^{+\infty} W(\theta, \bar{\theta}) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta = \int_{-\infty}^{\bar{\theta}} 2(\bar{\theta} - \theta) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta + \int_{\bar{\theta}}^{\infty} (\theta - \bar{\theta}) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta$$

becomes a minimum. As an easy calculation shows, the estimate obtained in this way is not the arithmetic mean.

As a second example we consider the family of variates $X(\theta)$ with the probability density $f(x, \theta)$ defined as follows:

$$f(x, \theta) = 1 \quad \text{if} \quad \theta - \tfrac{1}{2} \leq x \leq \theta + \tfrac{1}{2}$$

and

$$f(x, \theta) = 0 \text{ for all other values of } x.$$

If $E = (x_1, \cdots, x_n)$ denotes a sample point where x_1 denotes the smallest and x_n denotes the greatest value in the sample, then

$$p(E \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) = 1 \quad \text{if} \quad x_n - \tfrac{1}{2} \leq \theta \leq x_1 + \tfrac{1}{2}$$

and

$$p(E \mid \theta) = 0 \quad \text{for all other values of } \theta.$$

The classical method of maximum likelihood cannot be applied here, since $p(E \mid \theta)$ is maximum for every value θ for which $x_n - \frac{1}{2} \leq \theta \leq x_1 + \frac{1}{2}$. It is obvious that for a broad class of weight functions the conditions I–V are satisfied. The estimate $\bar{\theta}(E) = \dfrac{x_1 + x_n}{2}$, where x_1 denotes the smallest and x_n the greatest value in the sample, satisfies the condition VI. Hence if the weight function satisfies also the condition VII, the best estimate of θ is given by $\bar{\theta}(E) = \dfrac{x_1 + x_n}{2}$.

Let us now calculate the best estimate of θ if the weight function is given as follows:

$$W(\theta, \bar{\theta}) = \theta - \bar{\theta} \qquad \text{if} \quad \bar{\theta} \leq \theta$$

and

$$W(\theta, \bar{\theta}) = 2(\bar{\theta} - \theta) \quad \text{if} \quad \bar{\theta} > \theta.$$

In this case the conditions I–V are satisfied but not the condition VII. We have to calculate the integral $\psi(\bar{\theta}, E)$ given in (28), which reduces in this case to

$$\begin{aligned}\psi(\bar{\theta}, E) &= \int_{x_n - \frac{1}{2}}^{x_1 + \frac{1}{2}} W(\theta, \bar{\theta})\, d\theta = \int_{x_n - \frac{1}{2}}^{\bar{\theta}} 2(\bar{\theta} - \theta)\, d\theta + \int_{\bar{\theta}}^{x_1 + \frac{1}{2}} (\theta - \bar{\theta})\, d\theta \\ &= 1.5\bar{\theta}^2 - [(x_1 + \tfrac{1}{2}) + 2(x_n - \tfrac{1}{2})]\bar{\theta} + \tfrac{1}{2}(x_1 + \tfrac{1}{2})^2 + (x_n - \tfrac{1}{2})^2.\end{aligned}$$

This expression becomes a minimum for

$$\bar{\theta} = \frac{x_1 + 2x_n - \frac{1}{2}}{3}.$$

Hence the best estimate of θ is given by this expression.

11. **Miscellaneous remarks.** Assumptions 1–6 of paragraph 8 are sufficient but not necessary for the proof of the Theorems 1–3 (Theorems 4–6 have been deduced without Assumptions 1–6). They can be weakened in many respects. The assumption that the parameter space is bounded can be dropped if we impose certain conditions on the weight function $W(\theta, \bar{\theta})$ and the probability density $p(E \mid \theta)$. It is certainly not necessary to assume that $W(\theta, \bar{\theta})$ and $p(E \mid \theta)$ are everywhere continuous. It is however doubtful whether Theorems 1–3 remain valid in the form in which they are stated, if we admit discon-

tinuities in a set of measure zero without imposing any other restrictions. Also Assumptions 4–6 can in all probability be essentially weakened.

I should like to mention that Assumption 4 is not as restrictive as it would appear. Let us make this clear in the case that the parameter space is a one-dimensional interval $[a, b]$. If we assume that $W(\theta, \bar{\theta})$ is a polynomial of the second degree in $\bar{\theta}$ and the coefficient of $\bar{\theta}^2$ is positive for every θ, and if $p(E \mid \theta) > 0$ for every E and θ, the Assumption 4 can easily be proved. In fact,

$$\psi(\bar{\theta}, E) = \int_a^b W(\theta, \bar{\theta}) p(E \mid \theta)\, df(\theta) = A(E) + B(E)\bar{\theta} + C(E)\bar{\theta}^2.$$

Since the coefficient of $\bar{\theta}^2$ in $W(\theta, \bar{\theta})$ is positive and since $p(E \mid \theta) > 0$ for every E and θ, $C(E) > 0$ for every E and for any arbitrary distribution $f(\theta)$. From this fact follows easily that for every E there exists a value $\bar{\theta}(E)$ in the interval $[a, b]$ such that

$$\psi[\bar{\theta}(E), E] < \psi(\bar{\theta}, E)$$

for every $\bar{\theta}$ contained in $[a, b]$ and unequal to $\bar{\theta}(E)$. Hence Assumption 4 is proved.

Let us consider a system S of subsets of the parameter space Ω and the corresponding system H_S of hypotheses. The weight function $W(\theta, \omega)$ is defined for all points θ of Ω and for all elements ω of S and expresses the weight of the error committed by accepting H_ω when θ is true. If θ is an element of ω then $W(\theta, \omega)$ is of course equal to zero. Let us assume that $W(\theta, \omega)$ has the special form: $W(\theta, \omega) = 1$ if θ is not contained in ω, and $W(\theta, \omega) = 0$ if θ is an element of ω. It is obvious that in this case for any θ the value of the risk function $r(\theta)$ is equal to the probability of accepting a false hypothesis if θ is the true parameter point. Because of this fact the theory developed here has close relation to the theory of confidence intervals. Let us first make this clear for the case when the parameter space is one dimensional, that is to say θ is a real number.

In the theory of confidence intervals we estimate the unknown parameter θ by an interval $I(E)$ extending from $\theta_1(E)$ to $\theta_2(E)$ where $\theta_1(E)$ and $\theta_2(E)$ are certain functions of the sample point E. The interval $I(E)$ is defined in such a way that the following probability statement holds: If we perform an experiment, the probability that we shall obtain a sample point E such that $I(E)$ will cover the true parameter point θ, is equal to a given constant α (called confidence coefficient) and is independent of the value of θ. Let us consider a certain example of such an inference with the confidence coefficient α and denote by $I(E)$ the interval corresponding to E. We define a system S of intervals as follows: An interval I is an element of S if and only if there exists a sample point E for which $I(E) = I$. Consider the corresponding system H_S

of hypotheses and the weight function $W(\theta, I)$ defined for all values θ and all elements I of S as follows.

$$W(\theta, I) = 0 \quad \text{if} \quad \theta \text{ is a point of } I$$

$$W(\theta, I) = 1 \quad \text{if} \quad \theta \text{ is not contained in } I.$$

Denote by M_S a best system of regions of acceptance relative to the weight function defined above. Denote by $I'(E)$ the element of S which we accept according to M_S if E is the sample point. On account of the special form of the weight function, the risk is obviously equal to the probability of accepting a false interval. From the definition of the best system of regions it follows that for any θ the probability that $I'(E)$ will cover θ is greater than or equal to α. If the risk function is constant, that is to say, if the probability that $I'(E)$ will cover the true parameter point θ is independent of the value of θ, then the intervals $I'(E)$ are confidence intervals corresponding to a confidence coefficient $\alpha' \geq \alpha$.

Similar observations can be made if the parameter space is k-dimensional ($k > 1$) that is to say, θ is a system of k numbers $\theta^{(1)}, \cdots, \theta^{(k)}$ An important case is that when we have to estimate only one of the components, say $\theta^{(1)}$, by an interval. As the investigations of W. Feller[8] have shown, confidence intervals in such cases do not exist always. That is to say, it is not always possible to determine $I(E)$ such that the probability that $I(E)$ will cover $\theta^{(1)}$ is equal to a given constant α independently of the values of $\theta^{(1)}, \cdots, \theta^{(k)}$. It is of great interest to know under what conditions confidence intervals exist. I should like to mention that a further development of the theory given in paragraph 8 may contribute much to the solution of this problem. In order to make this clear, let us consider a system S_1 of one-dimensional intervals. To each element I of S_1 let there correspond the subset ω of the k-dimensional parameter space Ω consisting of all points $\theta = (\theta^{(1)}, \cdots, \theta^{(k)})$ for which $\theta^{(1)}$ lies in I. Consider the system S of subsets ω of Ω corresponding to all elements of S_1 and the system H_S of hypotheses corresponding to S. The weight function is to be chosen as follows: $W(\theta, \omega) = 1$ if θ is not an element of ω and $W(\theta, \omega) = 0$ if θ is an element of ω. Consider a best system M_S of regions of acceptance and the corresponding risk function $r(\theta)$. On account of the special definition of $W(\theta, \omega)$, $r(\theta)$ is equal to the probability of accepting a false hypothesis if θ is the true parameter point. If the risk function $r(\theta)$ is identically equal to a constant α, we have confidence intervals corresponding to the confidence coefficient α. In order to see under what conditions the risk function is constant, we have to consider an equivalent problem (see paragraph 7) where the system of hypotheses is the system of all simple hypotheses and the weight function $W(\theta, \bar{\theta})$

[8] W. Feller, "Note on Regions Similar to the Sample Space," *Statistical Research Memoirs*, Vol. II, 1938.

is given according to formula (4). If $W(\theta, \bar{\theta})$ satisfies Assumptions 1-6, the risk function of the best estimate is constant. As we have mentioned, Assumptions 1-6 can be weakened. In order to get valuable results concerning the problem of the existence of confidence intervals, we have to weaken especially Assumption 2. In fact $W(\theta, \omega)$ takes only the values 1 and 0 and therefore $W(\theta, \bar{\theta})$ cannot be continuous.

Finally I should like to mention that the most stringent test as defined by Robert W. B. Jackson[9] is contained as special case in our general definition of the best system of regions of acceptance. Jackson considers a discontinuous parameter space Ω. Consider the problem of testing the hypothesis $\theta = \theta_0$ where θ_0 denotes a point of Ω. According to Jackson's definition we have the most stringent test if the critical region w_0 satisfies the condition: the maximum of the numbers A and B

$$A = P(E \in w \mid \theta_0), \quad B = \text{least upper bound of } P(E \in \bar{w} \mid \theta) \text{ formed for all } \theta \neq \theta_0,$$

becomes a minimum for $w = w_0$. $\bar{w}$ denotes the region complementary to w. It is easy to see that Jackson's definition of the most stringent test coincides with our definition of the best system of regions of acceptance in the following special case:

1) Ω is discontinuous 2) S consists only of two elements.

3) The weight function $W(\theta, \omega)$ is equal to 1 if θ is not contained in ω.

COLUMBIA UNIVERSITY.

---

[9] Robert W Jackson, "Testing Statistical Hypotheses," *Statistical Research Memoirs*, Vol. I, 1936.

THE DISTRIBUTION OF THE MULTIPLE CORRELATION COEFFICIENT IN PERIODOGRAM ANALYSIS

By D. M. Starkey

1. **Geometrical interpretation of the problem.** We begin with a summary of some recent work by Hotelling, in a form relevant to this particular problem.[1] He suggests that the general question of finding the distribution of the multiple correlation coefficient corresponding to a fitted regression of y upon x may be solved by evaluating definite integrals corresponding to invariants of certain curves, surfaces, etc. For the purposes of illustration we may consider the case of fitting the relation

$$Y = a + bf(x, k, \epsilon)$$

where f is an arbitrary function, and a, b, k, ϵ are constants, to the observations y, where we are given n values of $y, y_1, y_2, \cdots, y_n$ and the corresponding values of $x, x_1, \cdots, x_n$. We shall postulate that the y's are independent and normally distributed about a certain mean and that the regression may be fitted by means of the principle of least squares.

We must minimize the sum of squares

$$\sum_{\alpha=1}^{\alpha=n} (y_\alpha - Y_\alpha)^2 = \sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)]^2$$

and hence we differentiate with respect to a, obtaining the first condition for a minimum

$$\sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)] = 0.$$

In the following, all summations take place over a range $\alpha = 1$ to n. Then we have

$$a = \bar{y} - b\bar{f}$$

where

$$\bar{y} = \frac{\Sigma y_\alpha}{n}, \qquad \bar{f} = \frac{\Sigma f(x_\alpha, k, \epsilon)}{n}$$

Thus we minimize the sum of squares

$$\Sigma[(y_\alpha - \bar{y}) - b(f(x_\alpha, k, \epsilon) - \bar{f})]^2$$

[1] Harold Hotelling, "Tubes and spheres in n-spaces, and a class of statistical problem", *American Journal of Mathematics*, April, 1939

or, putting $y'_\alpha = y_\alpha - \bar{y}$

$$Y'_\alpha = Y_\alpha - \bar{Y} = bf(x_\alpha, k, \epsilon) - \bar{f}$$

we see that the quantity $\Sigma(y'_\alpha - Y'_\alpha)^2$ is to be minimized.

Geometrically we may regard the set of values $(y_1, \cdots, y_n)$ as defining a point in n-space, and $(Y_1, \cdots, Y_n)$ will also represent a point in n-space on the 4-dimensional surface which may be obtained by eliminating a, b, k, ϵ from the relations $Y = a + bf(x, k, \epsilon)$. The points $(y'_1, \cdots, y'_n)$ and $(Y'_1, \cdots, Y'_n)$ represent the orthogonal projections of $(y_1, \cdots, y_n)$ and $(Y_1, \cdots, Y_n)$ on the plane $\Sigma y_\alpha = 0$. Hence we have to minimize the distance between these projections, noticing that $(Y'_1, \cdots, Y'_n)$ now lies on the 3-dimensional projection of the surface on which $(Y_1, \cdots, Y_n)$ lies. The multiple correlation between the observed and fitted values is defined as

$$R = \frac{\Sigma(y_\alpha - \bar{y})(Y_\alpha - \bar{Y})}{\sqrt{\Sigma(y_\alpha - \bar{y})^2 \Sigma(Y_\alpha - \bar{Y})^2}} = \frac{\Sigma y'_\alpha Y'_\alpha}{\sqrt{\Sigma y'^2_\alpha \Sigma Y'^2_\alpha}}$$

and this is equal to $\cos\theta$, where θ is the angle between the lines joining the origin to the points $(y'_1, \cdots, y'_n)$ and $(Y'_1, \cdots, Y'_n)$. For the purpose of evaluating R we may thus consider the projections of these points on the unit sphere in $\Sigma y_\alpha = 0$ with centre the origin, these being

$$\left(\frac{y'_1}{\sqrt{\Sigma y'^2_\alpha}}, \cdots, \frac{y'_n}{\sqrt{\Sigma y'^2_\alpha}}\right) \quad \text{and} \quad \left(\frac{Y'_1}{\sqrt{\Sigma Y'^2_\alpha}}, \cdots, \frac{Y'_n}{\sqrt{\Sigma Y'^2_\alpha}}\right),$$

As by hypothesis the distribution of y has spherical symmetry about some point on the line $y_1 = y_2 = \cdots = y_n$, then the distribution of y' has spherical symmetry about the origin, and the probability distribution of the projection of y' on the unit sphere is uniform. The projection of Y' lies on a 2-dimensional surface on the $(n-2)$-dimensional sphere, and for a given Y' the probability that R is as great or greater than $\cos\theta$ is proportional to the volume of the sphere in the $(n-2)$-dimensional spherical space with centre Y' and geodesic radius θ, so that the total probability that R lies between $\cos\theta$ and 1 is equal to the ratio of the "area" of the portion of the unit sphere included by the envelope of these geodesic spheres to the "area" of the unit sphere. This envelope is that part of the unit sphere in $\Sigma y_\alpha = 0$ which is at a geodesic distance θ from the 2-dimensional surface on which the projection of Y' lies, termed a "tube" by Hotelling.

For very small values of θ it may be assumed that this ratio is equal to the area of the two-dimensional surface on which Y' lies, multiplied by a fixed multiple of θ^{n-4}. This is fairly evident intuitively, but has recently been substantiated by some results of Weyl[2] who shows that this is correct for small values of θ, and indicates a series from which could be derived a series of ascend-

[2] H. Weyl, "On the volume of tubes," *American Journal of Mathematics*, April, 1939

ing powers of θ^2 by which successive adjustments could be made for larger values of θ. The coefficients in this series are finite invariants of the surface in which we are working. If we accept the first approximation we must consider the question of the extent of the surface, which depends on the range of values of the parameters k, ϵ. The range which is eventually chosen depends on the needs of the practical statistician, while keeping in view the mathematical possibilities of effecting a solution. In the following work we consider in particular the case of periodogram analysis by putting $f(x, k, \epsilon) = \cos(kx + \epsilon)$.

2. **The case of periodogram analysis.** With the notation of the preceding paragraph, we fit

$$Y_\alpha = a + b \cos(kx_\alpha + \epsilon)$$

to data (x_α, y_α) $\alpha = 1, 2, \cdots, n$.

We shall assume that the variate x is a measurement of time or some other quantity for which the measurements are made at equal intervals, taken as unity for convenience, so that

$$x_1 = 0, \qquad x_2 = 1, \cdots, x_n = n - 1.$$

Now we shall see later that we are interested in values of k such that $0 < k < 2\pi$. For this range

$$\bar{f} = \frac{\Sigma \cos(kx_\alpha + \epsilon)}{n}$$

$$= \frac{\sin(\frac{1}{2}nk) \cos[\epsilon + \frac{1}{2}k(n-1)]}{n \sin(\frac{1}{2}k)}.$$

Hence, if Y'' represents the projection of Y' on the unit sphere

$$Y''_\alpha = \lambda\left[\cos(kx_\alpha + \epsilon) - \frac{\sin(\frac{1}{2}nk) \cos[\epsilon + \frac{1}{2}k(n-1)]}{n \sin(\frac{1}{2}k)}\right]$$

where λ is to be determined so that

$$\Sigma Y''^2_\alpha = 1.$$

Now

$$\Sigma Y''^2 = \lambda^2\left[\Sigma \cos^2(kx_\alpha + \epsilon) - \frac{\sin^2(\frac{1}{2}nk) \cos^2[\epsilon + \frac{1}{2}k(n-1)]}{n \sin^2(\frac{1}{2}k)}\right]$$

and

$$\Sigma \cos^2(kx_\alpha + \epsilon) = \frac{1}{2}n + \frac{1}{2}\frac{\sin nk \cos[2\epsilon + k(n-1)]}{\sin k}$$

and hence

$$\lambda = \frac{1}{\sqrt{\frac{1}{2}n + \frac{1}{2}\frac{\sin nk \cos[2\epsilon + k(n-1)]}{\sin k} - \frac{\sin^2(\frac{1}{2}nk) \cos^2[\epsilon + \frac{1}{2}k(n-1)]}{n \sin^2(\frac{1}{2}k)}}}$$

the expression being continous at $k = \pi$.

Then
$$Y''_\alpha = \lambda(\cos(kx_\alpha + \epsilon) - \hat{f})$$
$$= \lambda \cos(kx_\alpha + \epsilon) + \xi \text{ say.}$$

Regarding k and ϵ as curvilinear coordinates of a point on the surface, we apply the formula
$$\sqrt{EG - F^2}\, dk\, d\epsilon$$
for the element of surface area, where
$$E = \Sigma\left(\frac{\partial Y''_\alpha}{\partial k}\right)^2, \qquad F = \Sigma \frac{\partial Y''_\alpha}{\partial k}\frac{\partial Y''_\alpha}{\partial \epsilon}, \qquad G = \Sigma\left(\frac{\partial Y''_\alpha}{\partial \epsilon}\right)^2.$$

In evaluating these summations, we shall need the following results: $\Sigma Y''_\alpha = 0$, $\Sigma Y''^2_\alpha = 1$, from which we obtain

(1)
$$\Sigma \cos(kx_\alpha + \epsilon) = \frac{-n\xi}{\lambda}$$

(2)
$$\Sigma \cos^2(kx_\alpha + \epsilon) = \frac{1 + n\xi^2}{\lambda^2}.$$

Differentiating these relations, we have

(3)
$$\Sigma x_\alpha \sin(kx_\alpha + \epsilon) = \frac{\partial}{\partial k}\left(\frac{n\xi}{\lambda}\right)$$
$$= \frac{n\xi_k}{\lambda} - \frac{n\xi\lambda_k}{\lambda^2}$$

(4)
$$\Sigma \sin(kx_\alpha + \epsilon) = \frac{\partial}{\partial \epsilon}\left(\frac{n\xi}{\lambda}\right)$$
$$= \frac{n\xi_\epsilon}{\lambda} - \frac{n\xi\lambda_\epsilon}{\lambda^2}$$

(5)
$$\Sigma x_\alpha^2 \sin^2(kx_\alpha + \epsilon) = \frac{1}{2}\Sigma x_\alpha^2 + \frac{1}{4}\frac{\partial^2}{\partial k^2}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{n(n-1)(2n-1)}{12} + \frac{1}{4}\left[\left(-\frac{2\lambda_{kk}}{\lambda^3} + \frac{6\lambda_k^2}{\lambda^4}\right)(n\xi^2 + 1)\right.$$
$$\left. - \frac{8\lambda_k n\xi\xi_k}{\lambda^3} + \frac{2n}{\lambda^2}(\xi\xi_{kk} + \xi_k^2)\right]$$

(6)
$$\Sigma x_\alpha \cos(kx_\alpha + \epsilon)\sin(kx_\alpha + \epsilon) = -\frac{1}{2}\frac{\partial}{\partial k}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{\lambda_k(1 + n\xi^2)}{\lambda^3} - \frac{n\xi\xi_k}{\lambda^2}$$

$$(7)\qquad \Sigma \cos(kx_\alpha+\epsilon)\sin(kx_\alpha+\epsilon) = -\frac{1}{2}\frac{\partial}{\partial\epsilon}\left(\frac{1+n\xi^2}{\lambda^2}\right)$$
$$= \frac{\lambda_\epsilon}{\lambda^3}(1+n\xi^2) - \frac{n\xi\xi_\epsilon}{\lambda^2}$$

$$(8)\qquad \Sigma x_\alpha \sin^2(kx_\alpha+\epsilon) = \frac{1}{2}\Sigma x_\alpha + \frac{1}{4}\frac{\partial^2}{\partial k\partial\epsilon}\left(\frac{1+n\xi^2}{\lambda^2}\right)$$
$$= \frac{n(n-1)}{4} - \frac{1}{2}\frac{\lambda_{\epsilon k}}{\lambda^3}(1+n\xi^2) + \frac{3}{2}\frac{\lambda_\epsilon\lambda_k}{\lambda^4}(1+n\xi^2) - \frac{n\lambda_\epsilon\xi\xi_k}{\lambda^3}$$
$$+ \frac{n\xi_k\xi_\epsilon}{2\lambda^2} + \frac{n\xi\xi_{\epsilon k}}{2\lambda^2} - \frac{n\xi\xi_\epsilon\lambda_k}{\lambda^3}$$

Now

$$\frac{\partial Y''_\alpha}{\partial k} = \lambda_k\cos(kx_\alpha+\epsilon) - \lambda x_\alpha\sin(kx_\alpha+\epsilon) + \xi_k$$

and

$$\frac{\partial Y''_\alpha}{\partial \epsilon} = \lambda_\epsilon\cos(kx_\alpha+\epsilon) - \lambda\sin(kx_\alpha+\epsilon) + \xi_\epsilon$$

so that with the above definitions of E, F, G we obtain

$$E = \lambda_k^2\Sigma\cos^2(kx_\alpha+\epsilon) + \lambda^2\Sigma\sin^2(kx_\alpha+\epsilon) + n\xi_k^2 - 2\lambda\lambda_k\Sigma x_\alpha\cos(kx_\alpha+\epsilon)$$
$$\cdot\sin(kx_\alpha+\epsilon) - 2\lambda\xi_k\Sigma x_\alpha\sin(kx_\alpha+\epsilon) + 2\lambda_k\xi_k\Sigma\cos(kx_\alpha+\epsilon)$$
$$= -\frac{\lambda_{kk}}{2\lambda} + \frac{\lambda_k^2}{2\lambda^2} + \frac{n(n-1)(2n-1)}{12}\lambda^2 - \tfrac{1}{2}n\lambda^2\bar{f}_k^2 + \tfrac{1}{2}n\bar{f}^2\lambda^2\bar{f}_{kk}$$

$$F = \lambda_\epsilon\lambda_k\Sigma\cos^2(kx_\alpha+\epsilon) + \lambda^2\Sigma x_\alpha\sin^2(kx_\alpha+\epsilon) + n\xi_\epsilon\xi_k$$
$$-\lambda\lambda_k\Sigma\sin(kx_\alpha+\epsilon)\cos(kx_\alpha+\epsilon) + \xi_\epsilon\lambda_k\Sigma\cos(kx_\alpha+\epsilon) - \lambda\xi_\epsilon\Sigma x_\alpha\sin(kx_\alpha+\epsilon)$$
$$+\lambda_\epsilon\xi_k\Sigma\cos(kx_\alpha+\epsilon) - \lambda\lambda_\epsilon\Sigma x_\alpha\sin(kx_\alpha+\epsilon)\cos(kx_\alpha+\epsilon) - \lambda\xi_k\Sigma\sin(kx_\alpha+\epsilon)$$
$$= \lambda_\epsilon\lambda_k\left(\frac{1}{2\lambda^2}\right) - \frac{\lambda_{k\epsilon}}{2\lambda} + \frac{\lambda^2 n(n-1)}{4} + \frac{n\bar{f}\bar{f}_{k\epsilon}}{2}\lambda^2 - \frac{n\lambda^2\bar{f}_k\bar{f}_\epsilon}{2}$$

$$G = \lambda_\epsilon^2\Sigma\cos^2(kx_\alpha+\epsilon) + \lambda^2\Sigma\sin^2(kx_\alpha+\epsilon) + n\xi_\epsilon^2 - 2\lambda\xi_\epsilon\Sigma\sin(kx_\alpha+\epsilon)$$
$$- 2\lambda\lambda_\epsilon\Sigma\cos(kx_\alpha+\epsilon)\sin(kx_\alpha+\epsilon) + 2\lambda_\epsilon\xi_\epsilon\Sigma\cos(kx_\alpha+\epsilon)$$
$$= -\frac{\lambda_\epsilon^2}{\lambda^2} + n\lambda^2 - (n\bar{f}^2\lambda^2+1) - n\lambda^2\bar{f}_\epsilon^2,$$

after using the relation $\xi = -\bar{f}\lambda$ to eliminate ξ.

These relations give

$$EG - F^2 = \left(-\frac{\lambda_{kk}}{2\lambda} + \frac{\lambda_k^2}{2\lambda^2} + \frac{n-1}{12}n(2n-1)\lambda^2 - \frac{1}{2}n\lambda^2\bar{f}_k^2 + \frac{1}{2}n\bar{f}^2\lambda^2\bar{f}_{kk}\right)$$

$$\times\left(-\frac{\lambda_\epsilon^2}{\lambda^2} + n\lambda^2 - (n\bar{f}^2\lambda^2 + 1) - n\lambda^2\bar{f}_\epsilon^2\right) \tag{9}$$

$$-\left(\frac{\lambda_\epsilon\lambda_k}{2\lambda^2} - \frac{\lambda_{k\epsilon}}{2\lambda} + \frac{\lambda^2 n(n-1)}{4} + \frac{n\bar{f}\bar{f}_{k\epsilon}\lambda^2}{2} - \frac{n\lambda^2\bar{f}_k\bar{f}_\epsilon}{2}\right)^2.$$

The area of the surface on which Y'' lies is

$$\sqrt{EG - F^2}\, dk\, d\epsilon$$

over an appropriate range of values of k and ϵ, but it appears that this integral cannot be evaluated exactly. We shall obtain an approximation for large values of n, by obtaining approximations to λ, $\bar{f}$, and their derivatives, when n is large.

The range of periods, $\frac{2\pi}{k}$, will be considered to vary from quantities greater than one up to half the range, that is $\frac{1}{2}(n-1)$. This is chosen on the grounds that the intervals of time would be adjusted so that there would be no expectation of periods less than the interval, and that enough observations would be chosen to include at least two periods in the range. Although this supposes some a priori knowledge of the possible periods, it seems reasonable to expect that the experimenter would have at least a rough idea of the range of periods which might fit his data before attempting to fit a harmonic curve. This range gives a range of values of k from $4\pi/(n-1)$ to $2\pi(1-v)$ where v is arbitrarily small, but fixed. In all cases the epoch, ϵ, varies from 0 to 2π.

It is readily seen that the surface is traced out only once for this range of values of k, ϵ, so that the problem in its approximate form is reduced to that of the evaluation of the definite integral

$$\int_0^{2\pi}\int_{\frac{4\pi}{n-1}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk\, d\epsilon.$$

We shall obtain the approximations mentioned above, in the first place excluding from consideration values of k in the neighbourhood of 0, π, 2π, the integrals over small ranges including these values being obtained separately.

If k is not in the neighbourhood of 0, π, 2π, we note that

$$\frac{\sin(\frac{1}{2}nk)\cos[\epsilon + \frac{1}{2}k(n-1)]}{\sin(\frac{1}{2}k)}.$$

is a bounded function of k, the upper bound being independent of k, and at most equal to $|\operatorname{cosec}\frac{1}{2}k_1|$, where k_1 is the angle in the range considered nearest to 0, π, 2π. Similarly the upper bound of

$$\frac{\sin(nk)\cos[2\epsilon + k(n-1)]}{\sin k}$$

is at most $|\operatorname{cosec} k_1|$. Hence as n is increased, we may expand $\lambda/n^{\frac{1}{2}}$ in ascending powers of n^{-1}. For large n, therefore, $\lambda = O(n^{-\frac{1}{2}})$, and is approximately $(2/n)^{\frac{1}{2}}$. Since differentiation with respect to k introduces a multiplying factor n in some of the terms, it follows that this is compensated for by the factor λ^{-2} which occurs in the denominator of the derivative, and we may conclude that $\lambda_k = O(n^{-\frac{1}{2}})$. No such compensating factor n occurs in the numerator of λ_ϵ, and it is therefore of order $(n^{-\frac{1}{2}})$. It may readily be seen without actually evaluating the derivatives, which are very long and unwieldy expressions, that

$$\lambda_{kk} = O(n^{\frac{1}{2}}),\ \lambda_{\epsilon k} = O(n^{-\frac{1}{2}}),\ \bar{f}_\epsilon = O(n^{-1}),\ \bar{f}_k = O(1),$$

$$\bar{f}_{kk} = O(n),\ \bar{f}_{k\epsilon} = O(1),\ \bar{f} = O(n^{-1}).$$

We thus see that the term of highest order in $E = \dfrac{(n-1)n(2n-1)}{12}\lambda^2$.

The term of highest order in $G = n\lambda^2 - 1$.

The term of highest order in $F = \dfrac{n(n-1)}{4}\lambda^2$.

These are approximately constant for large n, and are equal to $n^2/3$, 1, $n/2$ to a first order of approximation. Hence

$$\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}.$$

The range for k may be broken up as follows:

(a) from $\dfrac{4\pi}{n-1}$ to $\dfrac{\alpha}{n^{\frac{1}{2}}}$, where α is a finite angle, independent of n.

(b) from $\dfrac{\alpha}{n^{\frac{1}{2}}}$ to $\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$

(c) from $\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $\pi + \dfrac{\alpha}{n^{\frac{1}{2}}}$

(d) from $\pi + \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $2\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$

(e) from $2\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $2\pi(1 - v)$.

The method of procedure will be to show that in ranges (a), (c), (e) the integrand is of order n, and that since the ranges in all three cases are of order $n^{-\frac{1}{2}}$, the values of the integrals in these ranges are $O(n^{\frac{1}{2}})$ which is negligible in comparison with the contributions from (b) and (d), which are $O(n)$.

In (a), $\dfrac{4\pi}{n-1} \leq k \leq \dfrac{\alpha}{n^{\frac{1}{2}}}$, we put $k = \dfrac{\alpha}{n^{1-\delta}}$, $\alpha = 4\pi$, and let δ range from p to $\frac{1}{2}$, where p is a positive quantity defined by the relation $(n-1) = n^{1-p}$. Then λ, $\bar{f}$, are of orders $n^{-\frac{1}{2}}$ and $n^{-\delta}$ respectively. For this range of values of δ, the orders of the derivatives are·

λ_k	λ_{kk}	λ_ϵ	$\lambda_{k\epsilon}$	$\bar{f}_k$	$\bar{f}_\epsilon$	$\bar{f}$
$n^{\frac{1}{2}-\delta}$	$n^{\frac{1}{2}-\delta}$	$n^{-\frac{1}{2}-\delta}$	$n^{\frac{1}{2}-\delta}$	$n^{1-\delta}$	$n^{-\delta}$	$n^{1-\delta}$

These being decreased, it follows that the order of $\sqrt{EG - F^2}$ is not increased for any positive δ, and $\sqrt{EG - F^2} = O(n)$ as before.

In (c), $\pi - \frac{\alpha}{n^{\frac{1}{2}}} \leq k \leq \pi + \frac{\alpha}{n^{\frac{1}{2}}}$, we put $k = \pi \pm \frac{\alpha}{n^{1-\delta}}$, according as $k \gtrless \pi$, and consider $0 \leq \delta \leq \frac{1}{2}$. The orders of the derivatives are as stated in (a) above for this range. The remainder of the range $\pi - \frac{\alpha}{n} < k < \pi + \frac{\alpha}{n}$ is such that the values of the derivatives are of orders as stated with $\delta = 0$, while $\lambda = O(n^{-1})$. Thus $\sqrt{EG - F^2} = O(n)$ throughout.

In (e), $2\pi - \frac{\alpha}{n^{\frac{1}{2}}} \leq k \leq 2\pi(1 - v)$, we put $k = 2\pi - \frac{\alpha}{n^{1-\delta}}$, and consider $0 \leq \delta \leq \frac{1}{2}$. In this range the orders of the derivatives are as in (a). In the remainder of the range, $2\pi - \frac{\alpha}{n} < k \leq 2\pi(1 - v)$, the orders of the derivatives are as in (a) with $\delta = 0$, so that $\sqrt{EG - F^2} = O(n)$.

As the ranges (b) and (d) are not independent of n, it remains to be shown that this fact does not affect the final result. We consider, therefore, $k = \frac{\alpha}{n^{1-\delta}}$ and $k = \pi - \frac{\alpha}{n^{1-\delta}}$ where $\frac{1}{2} < \delta < 1$, and since, as in (O) the second and third terms in the denominator of λ are $O(n^{1-\delta})$ and $O(n^{1-2\delta})$ or $O(n^{-1})$ respectively, $\lambda \sim 1 \Big/ \sqrt{\frac{n}{2}}$, while the derivatives have values as in case (a). Thus, in these ranges, $\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}$ throughout. Thus we may conclude in all cases that $\sqrt{EG - F^2} = O(n)$.

$$\text{The surface area} = \int_0^{2\pi} \left[\int_{\frac{4\pi}{n-1}}^{\frac{\alpha}{\sqrt{n}}} + \int_{\frac{\alpha}{\sqrt{n}}}^{\pi - \frac{\alpha}{\sqrt{n}}} + \int_{\pi - \frac{\alpha}{\sqrt{n}}}^{\pi + \frac{\alpha}{\sqrt{n}}} + \int_{\pi + \frac{\alpha}{\sqrt{n}}}^{2\pi - \frac{\alpha}{\sqrt{n}}} + \int_{2\pi - \frac{\alpha}{\sqrt{n}}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk \right] d\epsilon$$

$$= \int_0^{2\pi} \left[\int_{\frac{\alpha}{\sqrt{n}}}^{\pi - \frac{\alpha}{\sqrt{n}}} + \int_{\pi + \frac{\alpha}{\sqrt{n}}}^{2\pi - \frac{\alpha}{\sqrt{n}}} \sqrt{EG - F^2}\, dk \right] d\epsilon$$

$$+ \int_0^{2\pi} \left[\int_{\frac{4\pi}{n-1}}^{\frac{\alpha}{\sqrt{n}}} + \int_{\pi - \frac{\alpha}{\sqrt{n}}}^{\pi + \frac{\alpha}{\sqrt{n}}} + \int_{2\pi - \frac{\alpha}{\sqrt{n}}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk \right] d\epsilon.$$

In the first two ranges, $\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}$

In the last three ranges, $\sqrt{EG - F^2} = O(n)$ and therefore the integral $= O(n^{\frac{1}{2}})$.

Thus the area is equal to

$$(10)\quad \frac{n}{\sqrt{12}}\,2\pi\left(2\pi - \frac{4\alpha}{n^{\frac{1}{2}}}\right) + \text{terms of lower order} = \frac{4\pi^2}{\sqrt{12}}\cdot n = \frac{2\pi^2}{3}\sqrt{3}\,n$$

In the case of fitting a linear regression with 3 independent variates, the distribution of R is well known to be

$$\frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma(\frac{3}{2})\Gamma[\frac{1}{2}(n-4)]}\,(R^2)^{\frac{1}{2}}(1-R^2)^{\frac{1}{2}(n-6)}\,d(R^2).$$

It may readily be seen by a repetition of the argument used in the first paragraph that this expression could be derived by considering the volume of a tube in spherical space of $(n-2)$ dimensions, in which the base surface is a 2-dimensional unit sphere of area 4π. We are assuming that the first approximation to the volume of a tube is equal to the area of the surface multiplied by a fixed function of θ. If, therefore, we divide this expression by 4π, and take R sufficiently close to 1, or $\theta = \cos^{-1}R$ sufficiently close to zero, we shall obtain the expression by which to multiply the surface area, (10), in order to obtain the first approximation to the frequency function of R.

Using Stirling's approximation, we have

$$\Gamma[\tfrac{1}{2}(n-1)] \sim \sqrt{2\pi}\,e^{-\frac{1}{2}(n-1)}[\tfrac{1}{2}(n-1)]^{\frac{1}{2}(n-1)-\frac{1}{2}}$$

and

$$\Gamma[\tfrac{1}{2}(n-4)] \sim \sqrt{2\pi}\,e^{-\frac{1}{2}(n-4)}[\tfrac{1}{2}(n-4)]^{\frac{1}{2}(n-4)-\frac{1}{2}}.$$

The ratio of these $= e^{-\frac{1}{2}}2^{-\frac{1}{2}}\left(1+\frac{3}{n-4}\right)^{\frac{1}{2}(n-4)}(n-1)(n-4)^{\frac{1}{2}} \sim 2^{-\frac{1}{2}}n^{\frac{1}{2}}$.

Hence the multiplying constant is approximately $n^{\frac{1}{2}}/\sqrt{2\pi}$. Substituting $R = \cos\theta$ in the frequency function divided by this constant, we obtain $2\cos^2\theta\sin^{n-6}\theta\sin\theta\,d\theta$ giving $2\theta^{n-5}\,d\theta$ as the first approximation.

Hence the approximate frequency function for the quantity θ in the case of periodogram analysis is

$$\frac{n^{\frac{1}{2}}}{\sqrt{2\pi}}\,2\theta^{n-5}\,d\theta\,\frac{\frac{2\pi^2\sqrt{3n}}{3}}{4\pi} = 2^{-\frac{1}{2}}n^{\frac{1}{2}}\pi^{\frac{1}{2}}3^{-\frac{1}{2}}\theta^{n-5}\,d\theta.$$

Thus the first approximation to the probability that θ should be as great or greater is

$$2^{-\frac{1}{2}}n^{\frac{1}{2}}\pi^{\frac{1}{2}}3^{-\frac{1}{2}}\,\frac{\theta^{n-4}}{n-4}$$

or

$$n^{\frac{1}{2}}\left(\frac{\pi}{6}\right)^{\frac{1}{2}}\theta^{n-4}$$

approximately.

The approximations which have been introduced have been forced upon us by the limitations of the mathematical machinery involved. It must be admitted that these approximations are not those which the experimenter would choose, for the following obvious reason. If we are testing the null hypothesis that the population correlation is zero, for large values of n the sample correlation will approach its expectation value, namely zero, and we shall in general be interested in values of R which are small, and corresponding values of θ in the neighborhood of $\pi/2$. This situation is not provided for in this investigation. It may be, however, that there exists a large correlation in the population, and that owing to the large number in the sample the value of R calculated is near this value. Provided that this population correlation is sufficiently close to unity, the value of θ will be small enough to apply the distribution obtained above, and in such a case will enable us to reject the null hypothesis when the probability calculated from the distribution is sufficiently small.

UNIVERSITY OF LONDON.

ON THE APPLICATION OF THE Z-TEST TO RANDOMIZED BLOCKS

By M. D. McCarthy

1. **Introduction.** When a series of experiments is performed with the object of measuring some quantity, it is implicitly postulated that the quantity in question has a "true value," which is theoretically obtainable as the result of an infinite repetition of the experiment under the standard conditions. In certain experiments, especially those of physical and chemical science, the materials and the methods employed are subject to such accurate control by the experimenter that he can repeat his experiment again and again with the "essential" factors kept constant, and with biassed errors eliminated. This repetition gives a series of observations of the "true value" in question subject only to random errors. All that is needed, usually, to increase the accuracy of the estimate of the "true value" is to continue the repetition of the experiment. Not only does such a repetition make the estimate more exact but it also provides an estimate of the degree of accuracy present, permits a comparison between different quantities and makes it possible to test various hypotheses as to their relative values.

In many cases which arise, notably in biological and social science and in dealing with data provided by modern mass-production methods, it is a practical impossibility to repeat an experiment under the same essential conditions. The material available is definitely non-homogeneous with regard to at least some of the qualities influencing the results. In testing, for instance, a number of varieties of some plant, to find which gives the best yield, it is possible to guarantee, that to a high degree of accuracy all the varieties are cultivated alike. If a relatively small area is covered by the experimental plots, it can be said that all the varieties experience the same climatic conditions and it is not difficult to ensure that they are all treated alike as to measurement of produce and so on. It is, however, practically impossible to make the plots, on which the varieties are grown, homogeneous as regards fertility of the soil and, even if this were possible, it would partially defeat the purpose of the experiment which is to test the varieties over a certain limited range of soil types. In a similar way in many other fields of biological or social experiments a similar non-homogeneity of the experimental material exists.

In experimenting with homogeneous materials, where the conditions of the whole series of experiments are the same, the differences which occur between the theoretical "true value" and the observations are explained as being due to a multiplicity of causes outside the control of the experimenter and of such a nature that their incidence varies "randomly" from experiment to experiment. It is a fact that certain fundamental factors influencing the results are definitely non-random in their incidence which differentiates the experiments with

non-homogeneous material from the others and it is by *artificially* introducing randomization, as suggested by Fisher [1, 2, 3] that such experiments are made amenable to the usual error laws.

For convenience, in what follows, the word "variety" will be used when speaking of a single object of those under test, whether it actually be a variety of some plant, a manurial treatment, a method of feeding or anything else of the sort. For instance, if five varieties and three manurial treatments are being tested in the same experiment, a "variety" would be any one of the fifteen combinations of an actual variety under test with a manure. The word "plot" will be used for that portion of the non-homogeneous material which is required for the performance of an experiment on a single "variety," and the term "yield" will be applied to the value of the observed quantity obtained as the result of testing a "variety" on a single "plot." The plots are, of course, equalized with respect to "size," or whatever similar property would influence the test.

2. **Randomized Blocks.** Suppose that there are s varieties to be tested and that the necessary replication is attained by testing each variety on n separate plots. That the plots on which each variety is tested form a random sample of the material available is guaranteed by assigning each of the s varieties to n of the available ns plots *at random*, that is, as the result of a physical random experiment with cards, dice, or the like. This method of randomization may be so employed that no restrictions are put on the plots to which the varieties are assigned, or it may be further refined in different ways so that, while preserving the random nature of the assignment, certain restrictions may be placed on it. Such a method of randomization with restrictions is the method known as "randomized blocks."

The basic idea is that compact "blocks" of the non-homogeneous material are, probably, much more uniform than the material taken as a whole. Consequently, the material is first divided into n such "blocks," as compact and uniform as possible, each block containing s equal plots. Each of the s varieties under test is assigned to a single plot in every block and randomness is attained by making the assignment of the varieties to the plots in each block as the result of a separate random experiment Thus the n plots to which each variety is assigned do actually form a random sample of the non-homogeneous material with the restriction that to each plot of any variety corresponds a plot of any other variety from the same block.

3. **Mathematical Formulation.** $X_{jl(k)}$ denotes the "true yield" of the kth variety which would be obtained by testing it on the lth plot in the jth block. $k = 1, 2, \cdots, s$ denotes the number by which the variety is known, $l = 1, 2, \cdots, s$ the order-number of the plot in the block and $j = 1, 2, \cdots, n$ the number of the block. Following Neyman [4, p. 110] we define the "true yield," again with particular reference to agricultural experiments, as·

"Suppose that the experiment is repeated indefinitely without any change of

vegetative conditions or of arrangement so that the kth variety is always tested on the plot (j, l). The yields from this plot will form a population, say $\Pi_{jl(k)}$, and $X_{jl(k)}$ is defined as the mean of this population."

Thus, in any block, there are s^2 different possible populations with corresponding "true values," but in any single experiment on that block observations will be obtained from only s of the s^2 possible populations. To distinguish those populations for which an observation is available from those which are entirely hypothetical $X_{j(k)}$ will denote the "true yield," as already defined, of the kth variety *on the plot to which it has been assigned* in the jth block. Since this assignment has been carried out as the result of a random experiment the "true yield" is itself a *random variable*; $X_{j(k)}$ is randomly selected from the set of s possible values $X_{j1(k)}$, $X_{j2(k)}$, $\cdots$ $X_{js(k)}$.

Using the dot notation to denote the mean of a quantity taken over all values of the letter replaced by the dot, it is clear that

$$\begin{aligned} X_{jl(k)} &= X_{\ (k)} + [X_{j\ (k)} - X_{\cdot(k)}] + [X_{jl(k)} - X_{j\ (k)}] \\ &= X_{\cdot(k)} + B_{jk} + u_{jl(k)}, \end{aligned}$$

and

$$\begin{aligned} X_{j(k)} &= X_{\cdot\cdot(k)} + [X_{j\ (k)} - X_{\ (k)}] + [X_{j(k)} - X_{j\cdot(k)}] \\ &= X_{\cdot\cdot(k)} + B_{jk} + \eta_{jk}, \end{aligned}$$

where

$$B_{jk} = X_{j\ (k)} - X_{\cdot(k)}, \qquad u_{jl(k)} = X_{jl(k)} - X_{j\cdot(k)}$$

and

$$\eta_{jk} = X_{j(k)} - X_{j\cdot(k)}.$$

Obviously

$$\sum_{j=1}^{n} B_{jk} = 0 \quad \text{and} \quad \sum_{l=1}^{s} u_{jl(k)} = 0$$

from their definitions, while η_{jk} is a random variable, with zero expectation, selected from the sequence $u_{j1(k)}$, $u_{j2(k)}$ $\cdots$ $u_{js(k)}$. Neyman (loc. cit.) calls η_{jk}, thus defined, the "soil error" of the kth variety when tested on its assigned plot in the jth block. The actual yield observed when the kth variety is tested on its assigned plot in the jth block is x_{jk} and the difference $x_{jk} - X_{j(k)} = \epsilon_{jk}$ is termed the "technical error." Clearly

$$x_{jk} = X_{\cdot\cdot(k)} + B_{jk} + \eta_{jk} + \epsilon_{jk}. \tag{1}$$

Both "soil error" and "technical error" enter into any comparisons which may be made and it is well known that the major source of error in, for instance, agricultural experiments is that due to the heterogeneity of the soil. As regards the relative magnitudes of the two errors, that of course depends on the experiment in question, but Fisher [5] has stated that in an agricultural uniformity

trial (i.e. when the same variety is tested on all the plots) yields from plots of 1/40th of an acre frequently vary sufficiently among themselves, owing to soil heterogeneity, so as to give a standard deviation of ten per cent of the mean yield, while the inevitable random errors in treating the plots can be kept down to a much lower figure. By confining the randomization to a "block" of the material, which comprises only a relatively small compact portion of the whole material under test, the effects of soil heterogeneity may be much decreased. It appears, however, that it may very often be an unwarranted simplification to consider that the "true yield" of a variety is the same for all plots of a given block.

The two types of "error" are random variables of altogether different properties. Both have zero expectation and may be considered as independent of one another in the probability sense. It, therefore, appears reasonable to assume that ϵ_{jk} is independent both of the "technical error" in any other observation and of the η's. On the other hand η_{jk} is a random variable selected from the sequence

$$u_{j1(k)}\ ,\ u_{j2(k)}\ ,\ \cdots\ ,\ u_{js(k)} \tag{2}$$

and since, if η_{jk} has the value $u_{j1(k)}$ and η_{jm} is free to assume any one of the values $u_{j1(m)}\ ,\ u_{j2(m)}\ ,\ \cdots\ ,\ u_{js(m)}$ except $u_{j1(m)}$, it is clear that η_{jk} and η_{jm} are not independent. In the case of $\eta_{j'k}$ and $\eta_{j''m}$ where $j' \neq j''$, the random variables are drawn as the result of two separate random experiments from different sequences of the type (2). Obviously this means that the "soil errors" for different blocks are independent for either the same or different varieties. Writing E for the expected value, or the mean value in repeated experiments, since

$$\sum_{l=1}^{s} u_{jl(k)} = \sum_{l=1}^{s} u_{jl(m)} = 0,$$

the variance of η_{jk} is $\sigma^2_{\eta_{jk}}$ with

$$\sigma^2_{\eta_{jk}} = s^{-1} \cdot \sum_{l=1}^{s} u^2_{jl(k)} \tag{3}$$

and also

$$E[\eta_{jk}\eta_{jm}] = -\{s(s-1)\}^{-1} \sum_{l=1}^{s} u_{jl(k)}\, u_{jl(m)}\ . \tag{4}$$

Using (1), (3) and (4) it follows that

$$E[x_{jk}] = X_{\cdot(k)} + B_{jk} = a_{jk}\ , \tag{5}$$

say,

$$E[(x_{jk} - a_{jk})^2] = E[(\eta_{jk} + \epsilon_{jk})^2] = \sigma^2_{\eta_{jk}} + \sigma^2_{\epsilon}, \tag{6}$$

$$E[(x_{j'k} - a_{j'k})(x_{j''m} - a_{j''m})] = E[(\eta_{j'k} + \epsilon_{j'k})(\eta_{j''m} + \epsilon_{j''m})]$$
$$= E[\eta_{j'k}\eta_{j''m}].$$

The expectations of the various product terms on the right-hand sides of these equations vanish except in the case of the last one. If $j' \neq j''$ it too vanishes, whatever values of k and m, and it follows that the correlation of the observed yields of any two varieties, or of the same variety obtained from different blocks is zero. It is clear, however, that such is not the case when the yields are obtained from plots on the same block. Denoting by $\rho_{j(km)}$ the coefficient of correlation between x_{jk} and x_{jm} and using (4)

$$\rho_{j(km)} = \rho_{j(mk)} = -\frac{\sum_{l=1}^{s} u_{jl(k)} u_{jl(m)}}{s(s-1)\{\sigma_{\eta_j k}^2 + \sigma_{\epsilon_j}^2\}^{\frac{1}{2}}\{\sigma_{\eta_j m}^2 + \sigma_{\epsilon_j}^2\}^{\frac{1}{2}}}, \tag{7}$$

when $k \neq m$, while, of course, $\rho_{j(kk)} = 1$.

It may be noted that even when two sequences such as in (2) are identical the correlation $\rho_{j(km)}$ is not zero. In this case, when the varieties react in exactly the same way to variations in fertility within a block, $\sigma_{\eta_j k}^2 = \sigma_{\eta_j m}^2 = \sigma_{\eta_j}^2$, say, and

$$\rho_{j(km)} = -(s-1)^{-1}\{1 - \sigma_{\epsilon_j}^2/\sigma_{\eta_j}^2\}^{-1}. \tag{7a}$$

Then the coefficient of correlation is negative and depends only on the relative magnitude of the technical and soil errors for the block in question, and on the number of plots in the block. In a given block it is greatest in absolute magnitude when the technical error is zero, or at any rate negligible with respect to the soil error which, of course, is usually uncontrollable. In order to have zero correlation between the yields of every pair of varieties it must be assumed either that (a) there is such a complete lack of relationship between the ways in which the various varieties react to the differences of fertility within a block that for each pair of varieties k and m all the products such as $\sum_{l=1}^{s} u_{jl(k)} u_{jl(m)}$ vanish identically even though the u's themselves are not zero, an assumption that lacks plausibility, or else that (b) all members of each sequence of the type (2) are zero. This latter assumption means that no allowance whatsoever is needed for variations of fertility within a block. Once variation of fertility within a block is admitted it appears only reasonable that it should be taken into account and the effect of the resulting correlations on any test concerning the yields of different varieties examined.

Cramér has shown [6, 7] that if the sum of two independent random variables be normally distributed each variable must itself follow the normal law. Strictly, therefore, it cannot be correct to apply normal theory to the random variables x_{jk} in the mathematical model elaborated above, for, though ϵ_{jk} may readily be assumed normally distributed, η_{jk} can obviously take only a finite number, s, of values and, consequently, as its distribution cannot be normal, it is impossible that x_{jk} can be exactly normally distributed either. However, as a first approximation, taking into account the correlations, it will be assumed that the yields from any block form a set of single observations of the variables

in an s-variate normal distribution. Further, for the sake of simplicity, it will be assumed that the variances and covariances of the populations appropriate to the different blocks are the same. Dropping the distinguishing j's, the variances of the yields, as in (6) are defined by $\sigma_k^2 = \sigma_{\eta;k}^2 + \sigma_{\epsilon;}^2$ and ρ_{km} is written for $\rho_{j(km)}$ in (7). We define y_{jk} and A_{km} by

$$y_{jk} = x_{jk} - a_{jk} \tag{8}$$

and

$$A_{km} = \frac{\Delta_{km}}{2\sigma_k \sigma_m \Delta} = A_{mk} \tag{9}$$

where Δ is the s-rowed determinant $|\rho_{km}|$, symmetrical about its principal diagonal, Δ_{km} the cofactor of ρ_{km} in Δ and A is written for $|A_{km}|$ the determinant of the positive definite matrix $||A_{km}||$. Then since the interblock covariances are zero, the elementary probability law for the whole set of ns y's is given by

$$p\{y_{jk}\} = A^{\frac{1}{2}n} \pi^{-\frac{1}{2}ns} \exp\Big\{-\sum_j \sum_{k,m} A_{km} y_{jk} y_{jm}\Big\}. \tag{10}$$

It may be noted that j and l where they occur run through all integral values from 1 to n while k and m take values from 1 to s. A sign such as $\sum_{k,m}$ means that the summation is taken over all the pairs of values of k and m the term (m, k) being taken as distinct from the term (k, m) and including the terms in which $k = m$. $\sum_{k \neq m}$ implies a similar summation with the omission of terms in which $k = m$.

The distribution law (10), or similar one substituting the x's for y's from (8), takes into account also cases in which, though the correlations may be zero, the variances of the different variety yields differ.

4. The Z-Test. If $\{x_q\}$, $q = 1, 2, \cdots f_1$, is a set of f_1 mutually independent random variables each of which follows the same normal law with zero mean and variance σ_1^2, and if $u_1 = \sum_{q=1}^{f_1} x_q^2$, then the distribution law for u_1 is, $u_1 \geq 0$,

$$p(u) = \{\alpha^{\frac{1}{2}f}/\Gamma(\tfrac{1}{2}f)\} u^{\frac{1}{2}f-1} e^{-\alpha u^2} \tag{11}$$

with $f = f_1$ and $\alpha^{-1} = 2\sigma_1^2$. If also $u_2 = \sum_{r=1}^{f_2} y_r^2$, where $\{y_r\}$, $r = 1, 2, \cdots, f_2$, is another set of mutually independent random variables each of which is normally distributed with zero mean and variance σ_2^2, then the distribution law of u_2 is (11) with $\alpha^{-1} = 2\sigma_2^2$ and $f = f_2$. If, in addition to the independence of the variables within each set, there is also independence between the sets, then u_1 is independent of u_2 and the distributions of different functions of u_1

and u_2 used as criteria may be obtained. The one originally proposed in this connection was z, defined by,

$$z = \tfrac{1}{2} \log_e (f_1 u_2 / f_2 u_1) - \log_e (\sigma_2/\sigma_1)$$

and its distribution law is [8, 9, 10]

$$p(z) = \frac{2 f_1^{\frac{1}{2}f_1} f_2^{\frac{1}{2}f_2} \Gamma[\frac{1}{2}(f_1 + f_2)] e^{f_1 z}}{\Gamma(\frac{1}{2}f_1)\Gamma(\frac{1}{2}f_2)(f_1 + f_2 e^{2z})^{\frac{1}{2}(f_1+f_2)}}. \tag{12}$$

Any other single-valued, monotone function of u_2/u_1 would when $\sigma_1 = \sigma_2$, as a criterion, be equivalent to z. $F = e^{2z} = f_1 u_2 / f_2 u_1$, $v = u_2/u_1$ and $w = u_2/(u_1 + u_2)$ have been adopted as criteria and their distribution laws are readily deduced from (12). All these criteria are equivalent in providing control of "errors of the first kind" [11, 12], that is, the risk of rejecting a hypothesis tested when true. As usual the procedure is to select arbitrarily in advance a certain "level of significance," say $\epsilon = 0.05$, 0.01 etc., and, assuming the hypothesis tested is true, to determine the value of the criterion, say the value z_0 of z, such that

$$P\{z \geq z_0 \mid H\} = \int_{z_0}^{\infty} p(z)\, dz = \epsilon. \tag{13}$$

If the sample of observations gives a value of $z \geq z_0$ H is rejected, if $z < z_0$, H is accepted. It is merely a matter of convenience which of the criteria z, F, u or w is used and tables are available to facilitate numerical work. Tables for z and F are given by Fisher [2], Fisher and Yates [13] and Snedecor [14], while for w *Tables of the Incomplete Beta Function* [15] may be used. Though no tables are directly available for v it is the simplest to use in theoretical discussion and in subsequent sections it is its distribution law, and not that of z, which will be considered. The latter may, of course, be readily deduced.

Considering the distribution law (10) with y_{jk} replaced by $x_{jk} - a_{jk}$ when $\rho_{km} = 0$ and $\sigma_k = \sigma_m = \sigma$, i.e., all the observations are normal and independent with the same variance. Writing

$$u_1 = \sum_{j,k} (x_{jk} - x_{.k} - x_{j.} + x_{..})^2, \tag{14}$$

$$u_2 = \sum_{j,k} (x_{.k} - x_{..})^2 = n \sum_{k} (x_{.k} - x_{..})^2, \tag{15}$$

$$u_3 = \sum_{j,k} (x_{j.} - x_{..})^2 = s \sum_{j} (x_{j.} - x_{..})^2, \tag{16}$$

then it is readily seen that

$$u_1 + u_2 + u_3 = \sum_{j,k} (x_{jk} - x_{..})^2.$$

Now if a_{jk} may be put in the form $M + B_j + V_k$ with $\sum_j B_j = \sum_k V_k = 0$ then u_1 is distributed as in (11) with $f = (n-1)(s-1)$. If, in addition to the additive assumption, $V_k = 0$ for all values of k then u_2 follows the same law, independently of u_1, and with $f = s - 1$. Similarly if $B_j = 0$, for all values of j, u_3 has the same distribution law with $f = n - 1$. It may be shown [16] that if $a_{jk} = M$ for all values of j and k the three quantities u_1, u_2 and u_3 follow independently the law (11) with suitable values for f, and then the corresponding values of z follow the law (12).

Making the assumption of additivity for a_{jk}, of which, incidentally, the correctness or adequacy cannot be tested without more than one set of ns observations of the variables, the z-test may be used to determine whether or not there is a "block effect" or a "variety effect," i.e., whether $B_j = 0$ or $V_k = 0$ for all values of j and k. For instance to test the hypothesis $V_k = 0$, $k = 1, 2, \cdots, s$, $z = \frac{1}{2}\log_e\{(n-1)u_2/u_1\}$ is calculated from the observations and the hypothesis is rejected if $z \geq z_0$ where z_0 is found from Fisher's tables corresponding to a suitable value of ϵ in (13). Otherwise the hypothesis is accepted. This is the usual method of applying the z-test to randomized blocks.

The problem before us now is to consider what happens to such a test when $\sigma_k \neq \sigma_m$ and $\rho_{km} \neq 0$ in (10), and the hypotheses to be tested must be related to (1) and (5). As already stated this method of testing hypotheses controls, at a suitable level, the risk of rejecting the hypothesis when it is true. A complete examination of the application of any criterion as the test of a statistical hypothesis should involve, also, investigation of "errors of the second kind," i.e., the risk of accepting the hypothesis when some alternative is true. That is to say such an examination should involve a study of the "power function of the test" [17, 18, 19], and this would require a knowledge of the probability distribution of the criterion when the hypothesis tested is not true. In this paper, however, attention will be confined entirely to "errors of the first kind."

5. **Hypotheses Tested.** In order that

$$u_1 = \sum_{j,k} (y_{jk} - y_{\cdot k} - y_{j\cdot} + y_{\cdot\cdot})^2 \tag{14a}$$

and

$$u_2 = n \sum_k (y_{\cdot k} - y_{\cdot\cdot})^2 \tag{15a}$$

may be true it is sufficient that

$$a_{jk} - a_{\cdot k} - a_{j\cdot} + a_{\cdot\cdot} = X_{j\cdot(k)} - X_{\cdot\cdot(k)} - X_{j\cdot(\cdot)} + X_{\cdot\cdot(\cdot)}$$

and

$$a_{\cdot k} - a_{\cdot\cdot} = X_{\cdot\cdot(k)} - X_{\cdot\cdot(\cdot)}$$

must both be zero in every case. It has been suggested by Neyman [4] that it would be desirable to test the hypothesis that $X_{..(k)}$ is independent of k, i.e. that the average of the true yields over the whole field is the same for all varieties. He suggests that the variations in the responses of the different varieties within the field are relatively unimportant so that, while allowing for the effect of the variations in fertility within the field on the various distribution laws, it is the average over the whole field which should be tested. The functions u_1 and u_2 will not test this hypothesis for, in order that they may have the same expectation not only must $X_{(k)}$ be independent of k but also $X_{j(k)}$ must be independent of k for every j. Consequently one of the hypotheses tested here is that $X_{j(k)} = X_{j(\cdot)}$, and therefore, of course, $X_{..(k)} = X_{(\cdot)}$, for every j and k, i.e. that the mean of the true yields over all the blocks is the same for all varieties while, by using (10), we make allowance for the variations in fertility over each block and for the resultant correlations introduced. We shall not consider u_3 from (16) as we are interested only in the presence or absence of a "variety effect."

It appears that two other hypotheses lead to results which are particular cases of the above. If we test whether the true yield on every plot is the same for all varieties, i.e. that $X_{jl(k)}$ is independent of k, then, assuming the hypothesis tested is true, the varieties all react in the same way to the variations of fertility within each block and in (10) $\sigma_k = \sigma_m = \sigma$, say, while $\rho_{km} = \rho$. On the other hand if we neglect all the variations in fertility within each block all the correlations vanish and $\sigma_k = \sigma_m = \sigma_\epsilon$. The hypothesis tested then is that either $X_{jl(k)}$ or, what is the same thing, $X_{j\cdot(k)}$ is independent of k.

It does not appear that the assumption of normality need cause any difficulty. E. S. Pearson [20] has examined the effect of skewness on the parent populations and by carrying out sampling experiments has concluded that even with skew populations "... it seems probable that the more elaborate forms of analysis of variance are also of fairly wide application, provided that the number of degrees of freedom apportioned to the residual variation is not too small." A further investigation by Eden and Yates [21] was also designed to test the effect of skewness, but the negative result there obtained was to be expected owing to the amalgamation of the observations into groups. It appears that the effect of skewness in the original populations will not have very much effect on the distribution of z.

Welch has examined [22] Randomized Blocks and Latin Square experiments from the "randomization" point of view. In the case of randomized blocks, in terms of the notation used above, he has taken $\epsilon_{jk} = 0$ or, expressed in another way, he has assumed that the actual observed yield in any plot is the "true yield" on that plot of the particular variety tested on the plot. The hypothesis he is then testing is that $X_{jl(k)}$, or, what is the same thing for him $x_{jl(k)}$ is independent of k. Taking the $(s!)^n$ different ways in which the varieties may be tagged on to the different yields he has considered the $(s!)^n$ different values of what we have called, w and he has compared the finite discrete distri-

bution so obtained with that given by normal theory. Getting $E(w)$ and σ_w^2 from the finite distribution, he fitted a Pearson Type I curve in a number of examples and found that the 5 per cent and 1 per cent points in his fitted curves did not differ much from the corresponding points of the normal distribution of w. His theoretical discussion showed, however, that if there is too much discrepancy between the variancies in the different blocks the randomization test may seriously underestimate the significance of any differences between the varieties as compared with normal theory.

It was Neyman [4] who first pointed out that, when the variations of fertility within each block are taken into account, the correlations between the observed yields should be allowed for, and the method adopted here is a development of his point of view. A number of authors, however, while agreeing that such variations of fertility do occur, hold that this does not seriously affect the distribution of z.

6. **Distribution of u_1 and u_2.** As already stated, it is the distribution of $v = u_2/u_1$ which will be sought, not that of z, where u_1 and u_2 are defined by (14) and (15), or rather by (14a) and (15a), since the hypothesis tested is assumed true. Writing $i = \sqrt{-1}$, the characteristic function of the simultaneous distribution of u_1 and u_2, that is $E\,[\exp\{i(t_1u_1 + t_2u_2)\}]$, is found from (10).

From (14a) and (15a), by straightforward expansion, using the conventions already explained for $\sum_{k,m}$, $\sum_{k \neq m}$ etc., we get

$$u_1 = (ns)^{-1}\Big[(n-1)(s-1)\sum_{j,k} y_{jk}^2 - (n-1)\sum_{j}\sum_{k \neq m} y_{jk}y_{jm} - (s-1)\sum_{k}\sum_{j \neq l} y_{jk}y_{lk} + \sum_{j \neq l}\sum_{k \neq m} y_{jk}y_{lm}\Big],$$

$$u_2 = (ns)^{-1}\Big[(s-1)\Big\{\sum_{j,k} y_{jk}^2 + \sum_{k}\sum_{j \neq l} y_{jk}y_{lk}\Big\} - \sum_{j}\sum_{k \neq m} y_{jk}y_{jm} - \sum_{j \neq l}\sum_{k \neq m} y_{jk}y_{lm}\Big],$$

and using these expressions with (10) the characteristic function of u_1 and u_2 is

$$\varphi_{u_1,u_2}(t_1, t_2) = A^{\frac{1}{2}n}\pi^{-\frac{1}{2}ns}\int_{-\infty}^{\infty} p\{y_{jk}\}\cdot\exp\{i(t_1u_1 + t_2u_2)\}\,dY$$

$$= A^{\frac{1}{2}n}\pi^{-\frac{1}{2}ns}\int_{-\infty}^{\infty}\exp\Big\{\sum_{j,l}\sum_{k,m} B_{jk,\,lm}y_{jk}y_{lm}\Big\}\,dY$$

where $dY = \prod_{j,k} dy_{jk}$ and the integral is an ns-fold one taken over the whole space of these variables. $B_{jk,lm}$ is defined by

$$B_{jk,lm} = \delta_{jl}A_{km} - i(s\delta_{km} - 1)[t_1(n\delta_{jl} - 1) + t_2]/ns \tag{17}$$

where the δ's have the usual meaning being equal to 1 when the suffixes are the same and equal to zero when the suffixes are different. This integral, since the

real part of $B_{jk,lm}$ is positive definite, may readily be evaluated [23, 24] and gives

$$\varphi_{u_1,u_2}(t_1, t_2) = A^{\frac{1}{2}n}/B^{\frac{1}{2}} \tag{18}$$

B being the ns-rowed determinant $| B_{jk,lm} |$.

The determinant B may be written in the form

$$B = \begin{vmatrix} [P] & [Q] & \cdots & [Q] \\ [Q] & [P] & \cdots & [Q] \\ \cdots & \cdots & \cdots & \cdots \\ [Q] & [Q] & \cdots & [P] \end{vmatrix}$$

where $[P] = [p_{km}] = [B_{jk,jm}]$ and $[Q] = [q_{km}] = [B_{jk,lm}]$ and there are n^2 such arrays in B. This gives at once $B = | p_{km} + (n-1)q_{km} | \cdot | p_{km} - q_{km} |^{n-1}$, whence on substitution

$$B = | A_{km} - it_2(s\delta_{km} - 1)/s | \cdot | A_{km} - it_1(s\delta_{km} - 1)/s |^{n-1}. \tag{19}$$

The two determinants in (19) are identical, with t_1 and t_2 interchanged, and are readily reduced to symmetrical $(s-1)$-rowed determinants by: (a) Adding to the terms in the last row the corresponding terms in the other rows and repeating for columns, (b) Multiplying the terms in the last row successively by M_k/M $(k = 1, 2, 3, \cdots, s)$ and subtracting from the corresponding terms in each of the other rows, with

$$M_k = \sum_{m=1}^{s} A_{km} \quad \text{and} \quad M = \sum_{k=1}^{s} M_k = \sum_{k,m=1}^{s} A_{km}. \tag{20}$$

The following operations then reduce these $(s-1)$-rowed determinants to ones which are symmetrical and contain t's only in the diagonals: (i) To the terms in the last column add the corresponding terms in all the other columns and repeat for the rows, (ii) Multiply the terms in the last column by $(\sqrt{s}+1)^{-1}$ and add them to the corresponding terms in each of the other columns, repeat for the rows, (iii) From the terms in the last column subtract the sum of the corresponding terms in the other columns multiplied by $s^{-\frac{1}{2}}$, repeat for the rows, (iv) Divide the last row and the last column by $s^{-\frac{1}{2}}$. The determinant then becomes M/s. $| C - itI |$ where $|| C ||$ is the matrix $|| c_{km} ||$, I the unit matrix and

$$\begin{aligned} c_{km} = c_{mk} = A_{km} &- (\sqrt{s}+1)^{-1}(A_{ks} + A_{ms}) + (\sqrt{s}+1)^{-2}A_{ss} \\ &- M^{-1}[M_k - (\sqrt{s}+1)^{-1}M_s][M_m - (\sqrt{s}+1)^{-1}M_s]. \end{aligned} \tag{21}$$

It should be noted that henceforward k and m run through integral values from 1 to $s-1$ only unless the contrary is specifically stated.

Thus it follows that

$$\varphi_{u_1,u_2}(t_1, t_2) = (As/M)^{\frac{1}{2}n} | C - it_1 I |^{-\frac{1}{2}(n-1)} \cdot | C - it_2 I |^{-\frac{1}{2}}.$$

Putting $C = |\, c_{km} \,|$ and noting that $\varphi_{u_1, u_2}(t_1, t_2) = 1$ when $t_1 = t_2 = 0$ clearly $As/M = C$ and the characteristic function factors into the form $\varphi_{u_1}(t_1) \cdot \varphi_{u_2}(t_2)$, where

$$\varphi_{u_1}(t_1) = C^{\frac{1}{2}(n-1)} \,|\, C - it_1 I \,|^{-\frac{1}{2}(n-1)} \tag{22}$$

$$\varphi_{u_2}(t_2) = C^{\frac{1}{2}} \,|\, C - it_2 I \,|^{-\frac{1}{2}}. \tag{23}$$

This demonstrates that u_1 and u_2 are stochastically independent and that the correlations introduced by allowing for the variations in fertility within the blocks does not affect the independence already demonstrated [16, 25].

$\| C \|$ being a square positive definite matrix of rank $s - 1$, its characteristic equation $| C - \lambda I | = 0$ must have $s - 1$ real positive roots. It follows that $| C - itI |$ must factor into $s - 1$ factors of the type α-it where α is a real positive constant. Some or all of these factors may be equal and various combinations of factors of different multiplicity are possible depending on the value of s. Only two cases will be considered here: (a) when all the roots of the characteristic equation of $\| C \|$ are equal, and (b) when all the roots of the characteristic equation are unequal.

Suppose that all the roots of the characteristic equation are equal, say to α, then $| C - itI | = (\alpha - it)^{s-1}$ and $C = \alpha^{s-1}$ giving

$$\varphi_{u_1}(t_1) = \alpha^{\frac{1}{2}(n-1)(s-1)}(\alpha - it_1)^{-\frac{1}{2}(n-1)(s-1)}, \tag{24}$$

$$\varphi_{u_2}(t_2) = \alpha^{\frac{1}{2}(s-1)}(\alpha - it_2)^{-\frac{1}{2}(s-1)} \tag{25}$$

It is seen at once that u_1 and u_2 are distributed as in (11), $f_1 = (n - 1)(s - 1)$ and $f_2 = s - 1$, and thus z or v follow the usual distribution laws.

Clearly when the variations of fertility within each block are neglected and the hypothesis tested is that $X_{jl(k)}$, or $X_{j\,(k)}$, is independent of k, the roots of the characteristic equation are all equal. Then there is no correlation, $\sigma_k = \sigma_\epsilon$, $A_{kk} = (2\sigma^2)^{-1} = c_{kk} = \alpha$, $A_{km} = 0 = c_{km} (k \gtrless m)$ and the usual results are obvious.

On the other hand when allowing for the variations of fertility within a block while testing the hypothesis that $X_{il(k)}$ is independent of k, the variances and covariances are all equal, i.e. $\sigma_k^2 = \sigma_\eta^2 + \sigma_\epsilon^2 = \sigma^2$, $\rho_{kk} = 1$ and $\rho_{km} = \rho = -\sigma_\eta^2/\{(s - 1)(\sigma_\eta^2 + \sigma_\epsilon^2)\}$, $k \gtrless m$. This gives

$$\Delta_{km} = [\{1 + (s - 1)\rho\}\delta_{km} - \rho](1 - \rho)^{s-2}, \qquad \Delta = \{1 + (s - 1)\rho\}(1 - \rho)^{s-1},$$

$$A_{km} = \{1 + (s - 1)\delta_{km} - \rho\}/2\sigma^2\Delta, \qquad A = 1/(2\sigma^2)^s\Delta,$$

$$c_{km} = \delta_{km}\{2\sigma^2(1 - \rho)\}^{-1}, \qquad C = \{2\sigma^2(1 - \rho)\}^{-s+1},$$

where, as usual, $\delta_{km} = \begin{bmatrix} 1 & k = m \\ 0 & k \gtrless m \end{bmatrix}$ From this it follows the roots of the characteristic equation are all equal, $\alpha = \{2\sigma^2(1 - \rho)\}^{-1}$ in (24) and (25). Thus in this case also, the z-test or its equivalent gives exact control of errors of the first kind. There is, however, this difference that $u_1/(n - 1)(s - 1)$ and $u_2/(s - 1)$

are to be considered not as estimates of σ^2 but as estimates of $\sigma^2(1 - \rho) = (s-1)^{-1}\{s\sigma_\eta^2 + (s-1)\sigma_\epsilon^2\}$.

When $s = 2$, even though the variances differ, since there is only one root of the characteristic equation $\alpha = (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{-1}$ the characteristic functions are of the form (24) and (25). Consequently, in this case, $s = 2$, when only two varieties are tested for the hypothesis that their average "true yields" are the sa ie on each block then, even though the varieties may react in different ways to the fertility levels within the blocks, granting normality, the usual z-distribution applies. This, of course, includes the case when even though ρ may be zero the variances differ. $u_1/(n-1)$ and u_2 are to be considered as estimates of $\frac{1}{2}(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$.

Proceeding next to the case in which all the roots of the characteristic equation $|C - \lambda I| = 0$ are unequal, the roots are, say, $\alpha_1 < \alpha_2 < \cdots < \alpha_{s-1}$ where, of course, all these quantities are real and positive. This case will arise in testing the hypothesis that the yield for each variety is the same for every block, that

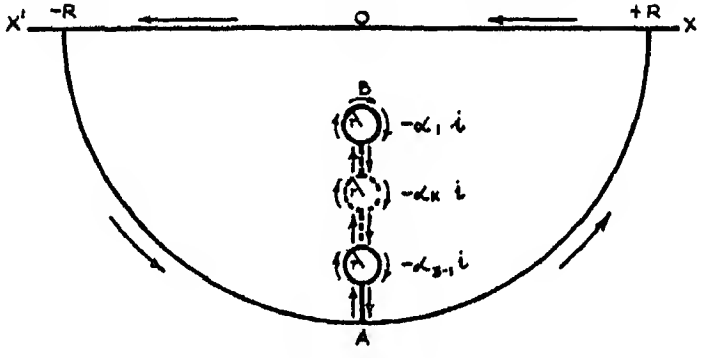


FIG. 1

$X_{j\cdot(k)} = X_{j\cdot(m)}$, while allowance is made for the different responses of the varieties to the differences in fertility within the blocks. The mathematical formulation would be the same even if there were no correlations but the variances were different for the different varieties. Then we have [26, 27, 28] for both u_1 and u_2 from (22) and (23)

$$p(u) = C^m(2\pi)^{-1}\int_{-\infty}^{\infty} e^{-iut}\Big[\prod_k (\alpha_k - it)\Big]^{-m} dt \tag{26}$$

with $m = \frac{1}{2}(n-1)$ for u_1 and $m = \frac{1}{2}$ for u_2.

Replace t by the complex variable z and integrate round the contour shown in Fig. 1. This contour consists of: (i) The real axis from $+R$ to $-R$, where $R > \alpha_{s-1}$, (ii) The quadrant $|z| = R$, $\pi \leq \arg z \leq 3\pi/2$, (iii) The imaginary axis from $A[-iR]$ to $B[-(\alpha_1 - r)i]$, cutting out the singularities by small semicircles of radius r, as shown, (iv) The imaginary axis from B to A, as in (iii), (v) The quadrant $|z| = R$, $3\pi/2 \leq \arg z \leq 2\pi$. Within this contour $f(z)$ is analytic

and hence the contour integral zero. It may also be readily seen that the integrals over the two quadrants tend to zero as R increases, and by examining the changes in the amplitudes of $(\alpha_k - iz)^{-m}$, $k = 1, 2, \cdots, s-1$, as the contour circles the points $-i\alpha_1, -i\alpha_2 \cdots$, it will be seen that the integrals over the straight lines between $(-i\alpha_2, -i\alpha_3)$, $(-i\alpha_4, -i\alpha_5)$, $\cdots$ cancel whether m be half an odd or half an even integer. Then

$$p(u) = C^m(2\pi)^{-1} \sum_r \int_{D_r} e^{-iuz} [\prod_k (\alpha_k - iz)]^{-m} dz. \tag{26a}$$

The contours D_r are those shown in Fig. 2 and consist of "dumb-bells" encircling the points $(-i\alpha_r, -i\alpha_{r+1})$, $r = 1, 3, 5, \cdots$, in the negative direction. If s is even, the last integral consists of only one half the "dumb-bell" extending to

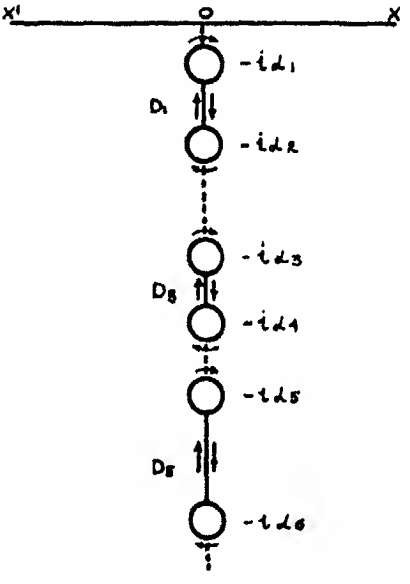


FIG. 2

$-i\infty$. It may also be noted that if n be odd and so m an integer, the other straight line integrals, those of the "dumb-bell" contours in question, also cancel and leave only the contributions of the small circles about, what are now, the poles.

Now we put $iz = w$ and $\Phi_r(w) = \prod_{k=1}^{s-1} (\alpha_k - w)^{-m}$ *omitting the terms containing* α_r *and* α_{r+1}, it follows that $\Phi_r(w)$ is analytic in and on a circle of centre $\frac{1}{2}(\alpha_r + \alpha_{r+1})$ which contains the points α_r and α_{r+1} but not α_{r-1} or α_{r+2}. Thus the function $\Phi_r(w)$ may in the interior of this circle be expanded as a uniformly convergent series in terms of $\frac{1}{2}(\alpha_r + \alpha_{r+1}) - w$ giving, $r = 1, 3, \cdots$,

$$\Phi_r(w) = \sum_{p=0}^{\infty} a_{rp} \{\tfrac{1}{2}(\alpha_r + \alpha_{r+1}) - w\}^p, \tag{27}$$

Since termwise integration is then permissible it is necessary to consider only integrals of the form

$$J_{rp} = i \int_{D_r'} \frac{\{\frac{1}{2}(\alpha_r + \alpha_{r+1}) - w\}^p e^{-uw} dw}{\{(\alpha_r - w)(\alpha_{r+1} - w)\}^m},$$

where D_r' is a contour similar to D_r but circling instead in a positive direction the points α_r and α_{r+1} on the real axis. We then have

$$p(u) = C^m (2\pi)^{-1} \sum_r \sum_{p=0}^{\infty} a_{rp} J_{rp}. \tag{26b}$$

Now if

$$J_r = i \int_{D_r'} \frac{e^{-uw} dw}{\{(\alpha_r - w)(\alpha_{r+1} - w)\}^m}$$

it is clear that J_{rp} is obtained by applying the operator $\left\{\frac{1}{2}(\alpha_r + \alpha_{r+1}) + \frac{\partial}{\partial u}\right\}$ p times to J_r. Now putting $w - \frac{1}{2}(\alpha_r + \alpha_{r+1}) = (\alpha_{r+1} - \alpha_r)t$, it follows that

$$J_r = \frac{ie^{-\frac{1}{2}u(\alpha_r + \alpha_{r+1})}}{e^{2m\pi i}\{\frac{1}{2}(\alpha_{r+1} - \alpha_r)\}^{2m-1}} \int_{\infty}^{(1+,-1+)} \frac{e^{-\frac{1}{2}ut(\alpha_{r+1} - \alpha_r)}}{(t^2 - 1)^m} dt$$

and this gives [29, p. 171],

$$J_r = \frac{\pi\Gamma(\frac{1}{2})u^{m-\frac{1}{2}} e^{-\frac{1}{2}u(\alpha_r + \alpha_{r+1})} I_{m-\frac{1}{2}}\{\frac{1}{2}u(\alpha_{r+1} - \alpha_r)\}}{2^{m-\frac{1}{2}}\Gamma(m)\{\frac{1}{2}(\alpha_{r+1} - \alpha_r)\}^{m-\frac{1}{2}}}.$$

$I_\mu(z)$ is the Bessel Function of purely imaginary argument defined, $-\pi < \arg z \leq \frac{1}{2}\pi$, by

$$I_\mu(z) = \sum_{r=0}^{\infty} \frac{(\frac{1}{2}z)^{\mu+2r}}{r!\Gamma(\mu + r + 1)}. \tag{28}$$

Hence it may be found that

$$J_{rp} = \frac{2\pi\Gamma(\frac{1}{2})}{\Gamma(m)(\alpha_{r+1} - \alpha_r)^{m-\frac{1}{2}}} e^{-\frac{1}{2}u(\alpha_r + \alpha_{r+1})} \frac{\partial^p}{\partial u^p}[u^{m-\frac{1}{2}} I_{m-\frac{1}{2}}\{\tfrac{1}{2}u(\alpha_{r+1} - \alpha_r)\}]$$

and this gives

$$p(u) = \frac{C^m\Gamma(\frac{1}{2})}{\Gamma(m)} \sum_r \frac{e^{-\frac{1}{2}u(\alpha_r + \alpha_{r+1})}}{(\alpha_{r+1} - \alpha_r)^{m-\frac{1}{2}}} \sum_{p=0}^{\infty} a_{rp} \frac{\partial^p}{\partial u^p}[u^{m-\frac{1}{2}} I_{m-\frac{1}{2}}\{\tfrac{1}{2}u(\alpha_{r+1} - \alpha_r)\}] \tag{29}$$

where a_{rp} is defined by (27) with $m = \frac{1}{2}(n-1)$ for u_1 and $m = \frac{1}{2}$ for u_2.

In the case $s = 3$ there is only one "dumb-bell" contour and $\Phi_1(w) = 1$, so that we get, for $u_1, u_2 \geq 0$.

$$p(u_1) = \frac{(\alpha_1\alpha_2)^{\frac{1}{2}(n-1)}\Gamma(\frac{1}{2})}{(\alpha_2 - \alpha_1)^{\frac{1}{2}(n-2)}\Gamma\{\frac{1}{2}(n-1)\}} e^{-\frac{1}{2}u_1(\alpha_1 + \alpha_2)} u_1^{\frac{1}{2}(n-2)} I_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}u_1(\alpha_2 - \alpha_1)\} \tag{30}$$

$$p(u_2) = (\alpha_1\alpha_2)^{\frac{1}{2}} e^{-\frac{1}{2}u_2(\alpha_1 + \alpha_2)} I_0\{\tfrac{1}{2}u_2(\alpha_2 - \alpha_1)\}. \tag{31}$$

It may be noted that if the series in (28) is substituted in (30) and (31) these distributions may be considered as the sum of an infinite number of χ^2-distributions all with the same σ^2 but with different degrees of freedom. It may also be noted that with $\alpha_1 = \alpha_2$ all the terms, except the first, vanish and thus a single χ^2-distribution is left.

When s is even the last contour is one from $+\infty$ circling α_{s-1} negatively. Using Hankel's integral for the Gamma-Function, putting $w - \alpha_{s-1} = \zeta/u$

$$I = i\int_{\infty}^{(\alpha_{s-1}+)} e^{-uw}(\alpha_{s-1} - w)^{-m}\,dw = ie^{-u\alpha_{s-1}}u^{m-1}\int_{\infty}^{(0+)} (-\zeta)^{-m}e^{-\zeta}\,d\zeta$$

$$= \frac{2\pi}{\Gamma(m)}\,e^{-u\alpha_{s-1}}u^{m-1}.$$

Denoting by D differentiation with respect to u under the sign of integration and by D^{-1} the corresponding integration from zero to u,

$$D^{-p}I = i\int_{\infty}^{(\alpha_{s-1}+)} (-w)^{-p}e^{-uw}(\alpha_{s-1} - w)^{-m}\,dw.$$

Then we can write

$$\prod_{k=1}^{s-2}(\alpha_k - w)^{-m} = (-w)^{-m(s-2)}\prod_{k=1}^{s-2}(1 - \alpha_k/w)^{-m}$$

$$= \sum_{p=0}^{\infty} a_{s-1,p}(-w)^{-p-m(s-2)}$$

the expansion being justifiable since $|\alpha_k/w| < 1$. Since $(s-2)m$ is an integer the additional term to be added to (29) to give $p(u)$ is, therefore,

$$C^m/\Gamma(m)\cdot\sum_{p=0}^{\infty} a_{s-1,p}D^{-p-m(s-2)}\{u^{m-1}e^{-u\alpha_{s-1}}\}. \tag{32}$$

7. **Distribution of $v = u_2/u_1$.** Though the distributions of u_1 and u_2 have been given in a rather complicated form for any value of s when the roots of $|C - \lambda I| = 0$ are all unequal, the distribution of v is given only for $s = 3$. In this case, since u_1 and u_2 are independent, from (30) and (31)

$$p(u_1, u_2) = \frac{(\alpha_1\alpha_2)^{\frac{1}{2}n}\Gamma(\frac{1}{2})}{(\alpha_2 - \alpha_1)^{\frac{1}{2}(n-2)}\Gamma\{\frac{1}{2}(n-1)\}}\,u_1^{\frac{1}{2}(n-2)}e^{-\frac{1}{2}(u_1+u_2)(\alpha_1+\alpha_2)}$$

$$I_0\{\tfrac{1}{2}u_2(\alpha_2 - \alpha_1)\}I_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}u_1(\alpha_2 - \alpha_1)\}$$

Now making the transformation $u_1 = u$, $u_2 = uv$ with $\dfrac{\partial(u_1, u_2)}{\partial(u, v)} = u$ and putting the exponential term

$$e^{-\frac{1}{2}u(1+v)(\alpha_1+\alpha_2)} = \frac{\{u(1+v)(\alpha_1+\alpha_2)\}^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}\,K_{\frac{1}{2}}\{\tfrac{1}{2}u(1+v)(\alpha_1+\alpha_2)\},$$

then on integrating with respect to u over the whole range of variation, from zero to infinity, we get

$$p(v) = \frac{(\alpha_1\alpha_2)^{\frac{1}{2}n}\{(1+v)(\alpha_1+\alpha_2)\}^{\frac{1}{2}}}{(\alpha_2-\alpha_1)^{\frac{1}{2}(n-2)}\Gamma\{\frac{1}{2}(n-1)\}}\int_0^\infty u^{\frac{1}{2}(n-1)}I_0\{\tfrac{1}{2}uv(\alpha_2-\alpha_1)\}$$
$$I_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}u(\alpha_2-\alpha_1)\}K_{\frac{1}{2}}\{\tfrac{1}{2}u(1+v)(\alpha_1+\alpha_2)\}\,du,$$

$K_j(z)$ being the modified Bessel Function of the second kind.

This integral is a particular case of one investigated by Bailey [30, 31] and it gives

$$(33)\quad p(v) = \frac{(n-1)(\alpha_1\alpha_2)^{\frac{1}{2}n}}{\{\frac{1}{2}(1+v)(\alpha_1+\alpha_2\}^n}F_4\{\tfrac{1}{2}(n+1),\tfrac{1}{2}n;1,\tfrac{1}{2}n;\beta^2v^2/(1+v)^2,\beta^2/(1+v)^2\}$$

where $\beta = (\alpha_2-\alpha_1)/(\alpha_2+\alpha_1)$ and F_4 is Appell's fourth hypergeometric function of two variables [32].

On performing a similar integration when $s > 3$, $p(v)$ may be obtained as a rather complicated series of terms similar to (33).

8. **Approximate Moments of the Distribution of v** As the distribution of v is complicated even in the simplest case of $s = 3$ it appears advisable to examine its moments even though only approximately. Writing $S_r = \sum_{k=1}^{s-1}\alpha_k^{-r}$ and putting

$$\prod_{k=1}^{s-1}(1-t/\alpha_k)^{-m} = \exp\left\{-m\sum_{k=1}^{s-1}\log(1-t/\alpha_k)\right\}$$
$$= \exp\left\{\sum_{r=1}^{\infty}mS_rt^r/r!\right\},$$

k_r being the r-th semi-invariant of u, we get

$$k_r = S_rm\cdot(r-1)!$$

Thence the first four moments of u about its mean are

$$\bar{u} = mS_1, \qquad \mu_2 = mS_2$$
$$\mu_3 = 2mS_3, \qquad \mu_4 = 3m\{2S_4 + mS_2^2\},$$

m being $\frac{1}{2}(n-1)$ in the case of u_1 and $\frac{1}{2}$ in the case of u_2.

Now to get approximate moments for v we write $\xi = (u_1-\bar{u}_1)/\bar{u}_1$ and $\eta = (u_2-\bar{u}_2)/\bar{u}_2$ and, expanding in terms of ξ and η, obtain

$$v = \bar{u}_2/\bar{u}_1\cdot\{1+\eta-\xi-\xi\eta+\xi^2+\xi^2\eta-\cdots\}$$

This gives, M_r being the r-th moment of v about the origin and writing $T_r = S_r/S_1^r = \sum_{k=1}^{s-1}\alpha_k^{-r}\Big/\left(\sum_{k=1}^{s-1}\alpha_k^{-1}\right)^r$,

$$M_1 = (n-1)^{-1}[1+2T_2(n-1)^{-1}-8T_3(n-1)^{-2}$$
$$+12\{4T_4+(n-1)T_2^2\}(n-1)^{-3}\cdots],$$

$$M_2 = (n-1)^{-2}(1+2T_2)[1+6T_2(n-1)^{-1}-32T_3(n-1)^{-2}$$
$$+60\{4T_4+(n-1)T_2^2\}(n-1)^{-3}\cdots],$$
$$M_3 = (n-1)^{-3}(1+6T_2+8T_3)[1+12T_2(n-1)^{-1}-80T_3(n-1)^{-2}$$
$$+180\{4T_4+(n-1)T_2^2\}(n-1)^{-3}\cdots],$$
$$M_4 = (n-1)^{-4}\{1+12T_2+32T_3+12(4T_4+T_2^2)\}[1+20T_2(n-1)^{-1}$$
$$-160T_3(n-1)^{-2}+420\{4T_4+(n-1)T_2^2\}(n-1)^{-3}\cdots].$$

The moments around the mean may readily be found if needed. If the α's are all equal

$$M_r' = \frac{\Gamma(\frac{1}{2}f_1 - r)\Gamma(\frac{1}{2}f_2 + r)}{\Gamma(\frac{1}{2}f_1)\Gamma(\frac{1}{2}f_2)}$$

from the known distribution of the ratio of two χ^2's with $f_2 = (s-1)$ and $f_1 = (n-1)(s-1)$ degrees of freedom respectively. Then developing M_r' as a series in terms of f_1^{-1} and f_2^{-1}

$$M_1' = (f_2/f_1)(1+2f_1^{-1}+4f_1^{-2}+\cdots),$$
$$M_2' = (f_2/f_1)^2(1+2f_2^{-1})(1+6f_1^{-1}+28f_1^{-2}+\cdots),$$
$$M_3' = (f_2/f_1)^3(1+6f_2^{-1}+8f_2^{-2})(1+12f_1^{-1}+100f_1^{-2}+\cdots),$$
$$M_4' = (f_2/f_1)^4(1+12f_2^{-1}+44f_2^{-2}+48f_2^{-3})(1+20f_1^{-1}+260f_1^{-2}+\cdots).$$

It is then easily seen that the difference between these moments and those of v when the α's are unequal is due to the deviation of T_r from $(s-1)^{r-1}$, the value it would have if the α's were all equal.

9. **Numerical Illustrations.** The distribution of v has been obtained in workable form only when $s = 3$ and, consequently, it is only that case that is considered here. In equation (33) the variable terms in the Appell function all contain $\beta = (\alpha_2 - \alpha_1)/(\alpha_2 + \alpha_1)$ and it is this fraction or, perhaps better, its square which measures, in a sense, the deviation of the distribution of v from the usual form. There are, therefore, two stages in this examination. It will first be investigated how β changes with the correlations and variances; and then the changes in the "levels of significance" due to differences in β will be examined.

Using (9), (20) and (21) it will be found that the equation $|C - \lambda I| = 0$ for $s = 3$ becomes $4p\lambda^2 - 4q\lambda + 3 = 0$ with

$$p = \sigma_2^2\sigma_3^2\Delta_{11} + \sigma_3^2\sigma_1^2\Delta_{22} + \sigma_1^2\sigma_2^2\Delta_{33} + 2\sigma_1\sigma_2\sigma_3(\sigma_1\Delta_{23} + \sigma_2\Delta_{31} + \sigma_3\Delta_{12}),$$
$$q = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 - \sigma_2\sigma_3\rho_{23} - \sigma_3\sigma_1\rho_{31} - \sigma_1\sigma_2\rho_{12},$$

where, of course, $\Delta = |\rho_{km}|$ and Δ_{km} is the cofactor of ρ_{km} in Δ. This equation may readily be solved giving α_1 and α_2.

Taking first the case of zero correlation and putting $\sigma_1^2 = k_1\sigma^2$, $\sigma_2^2 = k_2\sigma^2$ and

$\sigma_3^2 = k_3\sigma^2$ it will be found that while α_1 and α_2 depend on the k's and on σ^2 the fraction $\beta = (\alpha_2 - \alpha_1)/(\alpha_2 + \alpha_1)$ depends only on the k's. For different values of the k's the following table shows the values of β^2.

TABLE 1

Values of β^2 of different values of $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = k_1 : k_2 : k_3$. No correlation.

k_1	k_2	k_3	β^2
1	1	1	0
1.0	1.1	1.2	0.003
1.0	1.5	2.0	0.037
1	2	3	0.083
1	4	4	0.111
1	9	9	0.177
1	25	25	0.221
1	1	4	0.250
1	4	9	0.250
1	9	16	0.250
1	1	9	0.529
1	1	25	0.790
1	N	N	$(N-1)^2/(2N+1)^2$
1	1	N	$(N-1)^2/(N+2)^2$

It is clear that to get a considerable value of β^2, one standard deviation must be at least three times the other two. It also seems to produce a considerably larger value of β^2 to have one large k and two small k's than to have two large ones and one small, with the same order of magnitude of the ratio large to small. Furthermore, when the ratios of the σ^2 are $1:1:N$ the limit of β^2 as N increases is 1, while if the ratios are $1 : N : N$ the limit is 0.25.

Examining now the definition of ρ_{km}, omitting the j's in equation (7), we find that ρ_{km} can be written in the form

$$\rho_{km} = -r_{km}(s-1)^{-1}[(1 + \sigma_\epsilon^2/\sigma_{\eta_k}^2)(1 + \sigma_\epsilon^2/\sigma_{\eta_m}^2)]^{-\frac{1}{2}},$$

where

$$r_{km} = \frac{s^{-1}\sum_{l=1}^{n} u_{jl(k)}\, u_{jl(m)}}{\sigma_{\eta_k}\sigma_{\eta_m}}$$

and r_{km} is itself a coefficient of correlation, i.e., the correlation between the true yields. The second part of ρ_{km} depends on s and on the relative magnitudes of the soil error and of the technical error. Its maximum value, in the case of $s = 3$, is 0.5 which occurs when the technical error is small with respect to the soil error. If both types of error have the same variance then the second term is 0.25. There appears to be no data available which enables us to assign values to

r_{km}, so the method adopted is to choose some values of r_{km} which appear likely to affect seriously the value of β^2 and then to take the second factor equal 0.5. If the values of ρ_{km} are all equal and the variances are also equal the normal theory has been shown to apply, and hence these values are taken to differ considerably. Table 2 shows the effect on β^2 of taking different correlations with various values of $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$.

It is clear from the table that if there exist correlations of the order of magnitude of those assumed, they can cause the distribution of v to deviate considerably from that which arises on the usual theory. For instance, if the variances are equal the value of β^2 may be 0.444 a value it would attain if, with no correlations, one variance was seven times the other two. Taking the cases in which $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 4 : 9$ or $1 : 9 : 16$ the value of β^2 with no correlations is, in either case, 0.250 while with the correlations it may be as low as 0.008 or as high as 0.869.

TABLE 2

Values of β^2 for different values of the correlations and of $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$.

ρ_{12}	ρ_{13}	ρ_{23}	$\sigma_1^2:\sigma_2^2:\sigma_3^2$				
			1:1.1	1:4:9	1:9.16	1:25:25	1:1:25
0	0	0	0	0.250	0.250	0.221	0.790
−0.4	−0.1	0.2	0.099	0.083	0.075	0.059	0.721
0.2	−0.1	−0.4		0.523	0.549	0.543	0.851
−0.25	−0.25	0.2	0.074	0.132	0.104	0.056	0.766
0.2	−0.25	−0.25		0.423	0.429	0.402	0.843
0.4	0.2	−0.4	0.265	0.706	0.698	0.658	0.909
−0.4	0.2	0.4		0.028	0.019	0.020	0.690
0.4	0.4	−0.4	0.379	0.793	0.765	0.709	0.937
−0.4	0.4	0.4		0.016	0.016	0.038	0.673
0.4	0.4	−0.5	0.444	0.869	0.845	0.793	0.954
−0.5	0.4	0.4		0.008	0.010	0.042	0.654
0.4	0.1	−0.3	0.189	0.606	0.596	0.551	0.890
−0.3	0.1	0.4		0.055	0.035	0.012	0.721

On the other hand when β^2 is large, in the case of zero correlation, say $\beta^2 = 0.790$ when $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 1 : 25$, the correlations, as might be expected, appear to have less effect, the values of β^2 varying from 0.654 to 0.954. We may, therefore, conclude that if such correlations exist their effect on β^2, and therefore on the distribution of v, is certainly comparable with that of fairly large differences in the variances.

We now examine

$$P\{v > v_0 \,|\, \beta\} = \int_{v_0}^{\infty} p(v)\, dv$$

for different values of v_0 and β. Writing $p(v)$ in full from (33) and interchanging integration and summation we get

$$P\{v > v_0 \mid \beta\} = (n-1)(1-\beta^2)^{\frac{1}{2}n} \sum_{j,k=0}^{\infty} \frac{\{\frac{1}{2}(n+1)\}_{(j+k)}(\frac{1}{2}n+k)_{(j)}}{(j!)^2 k!} \beta^{2j+2k} \cdot \int_{v_0}^{\infty} v^{2j}(1+v)^{-n-2j-2k}\,dv.$$

Changing the variable to $x = (1+v)^{-1}$ the integral part becomes

$$\int_0^{x_0} x^{2k+n-2}(1+x)^{2j}\,dx = \frac{\Gamma(2j+1)\Gamma(2k+n-1)}{\Gamma(2j+2k+n)} I_{x_0}(2k+n-1, 2j+1),$$

in the notation usually employed [15]. Substitution gives

$$P\{v > v_0 \mid \beta\} = (1-\beta^2)^{\frac{1}{2}n} \sum_{j,k=0}^{\infty} \frac{(2j)!\{\frac{1}{2}(n-1)\}_{(k)}}{2^{2j}(j!)^2 k!} \beta^{2j+2k} I_{x_0}(2k+n-1, 2j+1).$$

Two sets of values of this expression were obtained, one for $n = 3$, and the other for $n = 6$, while β^2 was given the values 0.1, 0.2, 0.3, 0.4, and 0.5. The values of x_0 were chosen so as to cover the 1, 5 and 10 per cent significance levels. Table 3 shows these results.

TABLE 3

$P\{v > v_0/\beta\}$ for certain values of v_0 and β

(a) $n = 3$

x_0	v_0	Values of β^2					
		0.0	0.1	0.2	0.3	0 4	0.5
0.05	19	0.0025	0.003	0.003	0.004	0.004	0.005
0.10	9	0.010	0.011	0.013	0.014	0.016	0.018
0.15	$5\frac{2}{3}$	0.0225	0.025	0.027	0.030	0.034	0.037
0.20	4	0.040	0.043	0.048	0.051	0.056	0.061
0.30	$2\frac{1}{3}$	0.090	0.095	0.100	0.106	0.112	0.117
0.40	$1\frac{1}{2}$	0.160	0.165	0.170	0.175	0.181	0.187

(b) $n = 6$

x_0	v_0	Values of β^2					
		0.0	0.1	0 2	0.3	0 4	0.5
0.3	$2\frac{1}{3}$	0.002	0.003	0.004	0.005	0.006	0.007
0.4	$1\frac{1}{2}$	0.010	0.012	0.015	0.017	0.020	0.022
0.5	1	0.031	0.035	0.039	0.043	0.047	0.050
0.6	$\frac{2}{3}$	0.078	0.082	0.087	0.092	0.098	0.106
0.7	$\frac{3}{7}$	0.168	0.171	0.173	0.176	0.181	

The 1, 5, and 10 per cent levels of significance for x_0 were obtained in both cases by graphical interpolation and the corresponding values of v_0 then calculated. Table 4 shows clearly the changes in these significance levels. It must be remembered that values of β^2 considerably in excess of 0 5 may easily arise.

TABLE 4

Changes in the levels of significance $\beta^2 = 0$ *and* 0.5, $n = 3$ *and* 6.

		1 per cent		5 per cent.		10 per cent.	
		x_0	v_0	x_0	v_0	x_0	v_0
$n = 3$	$\beta^2 = 0$	0.10	9.00	0.22	3.47	0.32	2.16
	$\beta^2 = 0.5$	0.07	13.0	0.18	4.6	0.27	2.7
$n = 6$	$\beta^2 = 0$	0.40	1.51	0.55	0.82	0.63	0.58
	$\beta^2 = 0.5$	0.32	2.1	0.50	1.0	0.59	0.7

This work shows quite clearly that the effect of any correlation between the yields, such as that introduced by variations of fertility within a block, or of any difference between the yield variance of different varieties tends to cause a significant deviation to be recognised when, in fact, none exists. When the number of varieties tested is three, the variation in the levels of significance may be quite large.

10. **Conclusion.** The mathematical model appropriate to Randomized Block Experiments is examined and it is suggested that the use of the z-test, as ordinarily applied, is theoretically justifiable only when the variations in fertility within each block are negligible.

Correlations between the yields of the varieties, due to randomization in a limited set, are introduced when the differences in fertility within each block are allowed for.

It is suggested that, as a first approximation, a multinormal population may be used for the yields from a given block, the variances and correlations being assumed equal from block to block, though the means, of course, differ.

The simultaneous distribution of the usual sums of squares is found in this case, and these sums of squares are shown to be independently distributed as the sums of squares of $s - 1$ and $(n - 1)(s - 1)$ quantities from another multinormal population.

It is shown that the usual distribution results apply when the variances and correlations of all the varieties are equal as well, of course, as when the variances are equal and the correlations zero. It is also shown that the same is true when the number of varieties is two, though the variances may differ.

The distributions of the above sums of squares are obtained for all values of s,

the number of varieties, and the distribution of their ratio for $s = 3$. The method of obtaining the distribution of the ratio for $s > 3$ is also indicated.

The relative importance of the deviations from the usual distribution produced by differences in the variances and differences in the correlations is examined when $s = 3$, and it is found when the variances are all equal that the latter can produce deviations comparable to one variance being seven times the other two.

That the presence of the correlations or of non-equality of the variances causes a tendency for a significant difference to be found when none exists is clearly shown.

In conclusion, I must express my gratitude to Prof. J. Neyman, now of the University of California, for suggesting this problem to me, to Dr. R. C. Geary and to Prof. S. S. Wilks for valuable suggestions.

REFERENCES

[1] R. A. Fisher, *Jour. Minst. of Agric*, Vol. 33 (1926) pp. 503-513.
[2] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd (6th Ed. 1936).
[3] R. A Fisher, *The Design of Experiments*, Oliver & Boyd (1935)
[4] J. Neyman, *Suppl Jour Roy Stat. Soc*, Vol 2 (1935) pp. 107-180.
[5] R A Fisher, *Rothamsted Conferences* XIII (1931).
[6] H Cramér, *Math Zeitschrift*, Vol 4 (1936) pp. 405-414.
[7] H Cramér, *Random Variables and Probability Distributions*, Camb. Univ Press (1937).
[8] R. A. Fisher, *Proc. Intern. Math Congr, Toronto*, Vol. 2, (1924) pp. 805-813.
[9] R A. Fisher, *Econometrica*, Vol. 3 (1935) pp. 353-365.
[10] J O. Irwin, *Jour. Roy. Stat. Soc.*, Vol. 94 (1931) pp. 284-300.
[11] J Neyman and E. S Pearson, *Phil Trans. Roy. Soc. A.*, Vol. 231 (1933) pp. 289-337.
[12] J. Neyman and E. S. Pearson, *Proc. Camb. Phil. Soc.*, Vol. 29 (1933) 492-510.
[13] R A. Fisher and F. Yates, *Statistical Tables, for Biological, Agricultural, and Medical Research*, Oliver & Boyd (1938).
[14] G W. Snedecor, *Calculation and Interpretation of Analysis of Variance and Covariance*, Collegiate Press, Inc., Ames, Iowa, (1934).
[15] K. Pearson, *Tables of the Incomplete Beta Function*, Biometrika Office, London, (1934).
[16] J. O. Irwin, *Suppl Jour Roy. Stat. Soc.*, Vol. 1 (1934) pp. 236-251
[17] J. Neyman and E. S Pearson, *Stat. Research Mem.* Vol. 1 (1936) pp. 1-37
[18] J. Neyman, *Bull. de la Soc. Math. de France*, Vol. 63 (1935) pp. 246-266
[19] J Neyman, *Skandinavisk Aktuarietidskrift*, Vol. 20 (1931) pp. 149-199.
[20] E. S Pearson, *Biometrika*, Vol. 23 (1931) pp. 114-133.
[21] T. Eden and F. Yates, *Journ. Agric. Sci.*, Vol. 23 (1933) pp. 6-17
[22] B L. Welch, *Biometrika*, Vol. 29 (1937) pp. 21-52.
[23] J. Wishart and M S Bartlett, *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), pp. 260-270.
[24] A. C Aitken, *Quart. Journ. Maths.*, Vol. 2 (1931) pp. 130-135.
[25] W. G. Cochran, *Proc Camb. Phil. Soc*, Vol. 30 (1934) pp. 178-191.
[26] P. Levy, *Calcul des Probabilités*, Gauthier-Villars (1925).
[27] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill (1937)
[28] S S. Wilks, *Statistical Inference*, Edwards Brothers Inc. (1937).
[29] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Camb. Univ. Press (1922).
[30] W. N. Bailey, *Proc. Lond Math Soc.*, Vol. 40 (1935) pp. 37-48.
[31] W. N. Bailey, *Journ. Lond. Math. Soc*, Vol. 11 (1936) pp. 16-20.
[32] W. N. Bailey, *Generalized Hypergeometric Series*, Camb. Univ. Press (1937)

University College,
Cork, Ireland.

ON THE DISTRIBUTION OF ERRORS IN Nth TABULAR DIFFERENCES[1]

By Arnold N. Lowan and Jack Laderman

In the construction of mathematical tables, a frequent method of checking the computed values of the tabulated function is to apply a differencing test. This test consists of computing the tabular differences of some suitable order n and comparing them with the theoretical values of the differences computed to a higher degree of accuracy by an independent method. Whenever the absolute deviation of a tabular difference $\bar{\Delta}^n$ from the corresponding theoretical difference Δ^n exceeds some predetermined upper bound, the entries giving rise to the difference in question are investigated. Thus, in the computation of the functions $Si(x)$, $Ci(x)$, $Ei(x)$ and $Ei(-x)$, it was found desirable to check the final manuscript by comparing the tabular second differences with the values of the second differences computed to a higher degree of accuracy by an independent method.[2]

A study of the distribution of errors suggested the following problem: If we assume a rectangular distribution of the errors in the entries of a mathematical table, what is the distribution of errors in the n^{th} tabular differences?

For the sake of mathematical simplicity, it will be convenient to idealize the problem as follows: Consider $n + 1$ randon numbers $x_0, x_1, x_2, \cdots x_n$, drawn from any rectangular distribution. When arranged in a definite order, these $n + 1$ values give rise to an n^{th} difference Δ^n. If these $n + 1$ numbers are rounded to k decimal places, the new approximate values $\bar{x}_0, \bar{x}_1, \cdots \bar{x}_n$, will give rise to another nth difference $\bar{\Delta}^n$.

We shall investigate the distribution of the error $\Delta^n - \bar{\Delta}^n$.
The explicit expression for $\Delta^n - \bar{\Delta}^n$ is given by:

$$\Delta^n - \bar{\Delta}^n = C_0^n E_n - C_1^n E_{n-1} + C_2^n E_{n-2} - \cdots + (-1)^n C_n^n E_0$$

$$= w_0 + w_1 + w_2 + \cdots + w_n(\text{say})$$

where $E_i = x_i - \bar{x}_i$ and C_r^n are the binomial coefficients.

---

[1] The results reported in this paper were obtained in the course of work done by the Project for the Computation of Mathematical Tables. O. P. No. 765-97-3-10. Work Projects Administration, N. Y. C., operated under the sponsorship of Dr. Lyman J. Briggs, Director of the National Bureau of Standards. The authors wish to express their appreciation to the W. P. A. and to the Sponsor of this Project for permission to publish these results.

[2] The above functions were computed for $x = 0(0.0001)2.0000$ to 9 places of decimals and $x = 2(0.001)10.000$ to 9 decimals or significant figures. For the independent method of computation of the second differences, see article by A. N. Lowan in the *Bulletin of the American Mathematical Society*, August, 1939.

The distribution of any one of the E's is

$$f(E)\,dE = \begin{cases} 10^k\,dE & \text{if } -\tfrac{1}{2}10^{-k} \leq E \leq \tfrac{1}{2}10^{-k} \\ 0 & \text{elsewhere.} \end{cases}$$

The subsequent developments are based on the fundamental theorem which states that the characteristic function of the distribution of the sum of any number of random variables is the product of the characteristic functions of the distributions of the individual variables.[3] The characteristic function, $g(t)$, of $f(x)$ is defined as follows:

$$g(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx. \tag{1}$$

As is well known, the inversion of (1) is given by:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} g(t)\,dt. \tag{2}$$

It can be easily seen that the distribution of w_i is:

$$f(w_i)\,dw_i = \begin{cases} \dfrac{10^k}{C_i^n}\,dw_i, & \text{if } -\dfrac{10^{-k}C_i^n}{2} \leq w_i \leq \dfrac{10^{-k}C_i^n}{2} \\ 0 & \text{elsewhere.} \end{cases}$$

and its characteristic function is:

$$g_i(t) = \frac{\sin \tfrac{1}{2}(10^{-k}C_i^n t)}{\tfrac{1}{2}(10^{-k}C_i^n t)}.$$

On the basis of the theorem, above mentioned, the characteristic function of the distribution of $\Delta^n - \bar{\Delta}^n = y$ (say) is:

$$G(t) = \prod_{i=0}^{n} \frac{\sin \tfrac{1}{2}(10^{-k}C_i^n t)}{\tfrac{1}{2}(10^{-k}C_i^n t)}. \tag{3}$$

The desired frequency function, $F(y)$ is given by the inversion of

$$G(t) = \int_{-\infty}^{\infty} e^{itv} F(y)\,dy.$$

From (2) and (3), we get:

$$F(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itv} \prod_{i=0}^{n} \frac{\sin \tfrac{1}{2}(10^{-k}C_i^n t)}{\tfrac{1}{2}(10^{-k}C_i^n t)}\,dt$$

---

[3] See, for instance, Harald Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics No. 36 (1937) p. 36.

which may be written as:

$$F(y) = \frac{10^{k(n+1)}}{\pi \prod_{i=0}^{n} C_i^n} \int_{-\infty}^{\infty} \cos(2ty) \cdot \prod_{i=0}^{n} \sin(10^{-k} C_i^n t) \cdot \frac{dt}{t^{n+1}}. \tag{4}$$

The problem now reduces to the evaluation of the integral in (4). In the evaluation of this integral, it is convenient to consider even and odd values of n separately.

Case 1. When n is even

$$\prod_{i=0}^{n} \sin(10^{-k} C_i^n t) = \frac{(-1)^{\frac{1}{2}n}}{2^n} [P_{n+1} - P_n + P_{n-1} - \cdots + (-1)^{\frac{1}{2}n} P_{\frac{1}{2}n+1}]$$

where P_{n+1-r} denotes the sum of the sines of $n + 1 - r$ of the angles taken positively and the remaining r taken negatively, the negative angles being taken in every combination.[4] Thus $\cos(2ty) \prod_{i=0}^{n} \sin(10^{-k} C_i^n t)$ can be expressed as the sum of products of a sine by a cosine. By employing the identity $\sin A \cos B = \frac{1}{2}\{\sin(A + B) + \sin(A - B)\}$, each term can be written as the sum of two sines. Hence the integral under consideration can be written as the sum of integrals of the form

$$\int_{-\infty}^{\infty} \frac{\sin at}{t^{n+1}} dt.$$

Integrating by parts n times in succession, we obtain:

$$\int_{-\infty}^{\infty} \frac{\sin at}{t^{n+1}} dt = \frac{(-1)^{\frac{1}{2}n} a^n}{n!} \int_{-\infty}^{\infty} \frac{\sin at}{t} dt \tag{5}$$

But

$$\int_{-\infty}^{\infty} \frac{\sin at}{t} dt = \begin{cases} \pi & \text{for } a > 0 \\ -\pi & \text{for } a < 0. \end{cases}$$

(6) Therefore

$$\int_{-\infty}^{\infty} \frac{\sin at}{t^{n+1}} dt = \begin{cases} \dfrac{(-1)^{\frac{1}{2}n} a^n \pi}{n!} & \text{for } a > 0 \\ \dfrac{(-1)^{\frac{1}{2}n+1} a^n \pi}{n!} & \text{for } a < 0. \end{cases}$$

By use of (6) the integral in (4) can be readily evaluated.

Case 2. When n is odd.

$$\prod_{i=0}^{n} \sin 10^{-k} C_i^n t = \frac{(-1)^{\frac{1}{2}(n+1)}}{2^n} [Q_{n+1} - Q_n + Q_{n-1} - \cdots + (-1)^{\frac{1}{2}(n+1)} \cdot \tfrac{1}{2} Q_{\frac{1}{2}(n+1)}]$$

where Q_{n+1-r} denotes the sum of cosines of the sum of $n + 1 - r$ of the angles taken positively and the remaining r taken negatively, the negative angles being

[4] See E. W. Hobson, *Plane Trigonometry*, Seventh Edition (1928) pp. 50-51

taken in every combination. As in Case I, the integral in (4) can be expressed as the sum of integrals of the form:

$$\int_{-\infty}^{\infty} \frac{\cos at}{t^{n+1}}\, dt.$$

Integrating by parts, we obtain.

$$\int_{-\infty}^{\infty} \frac{\cos at}{t^{n+1}}\, dt = -\frac{a}{n} \int_{-\infty}^{\infty} \frac{\sin at}{t^{n}}\, dt.$$

The second member of this equation has been treated in Case I. It follows that:

$$\int_{-\infty}^{\infty} \frac{\cos at}{t^{n+1}}\, dt = \begin{cases} \dfrac{(-1)^{\frac{1}{2}(n+1)} a^n \pi}{n!} & \text{for } a > 0 \\ \dfrac{(-1)^{\frac{1}{2}(n-1)} a^n \pi}{n!} & \text{for } a < 0. \end{cases} \tag{7}$$

By means of the integrals, (6) and (7), $F(y)$ can be obtained for any n. The results for $n = 1$, 2, and 3 are given below:

$n = 1$

$$F(y) = \begin{cases} 10^{2k} y + 10^k & \text{for } -10^{-k} < y \leq 0 \\ -10^{2k} y + 10^k & \text{for } 0 \leq y < 10^{-k} \\ 0 & \text{elsewhere} \end{cases}$$

$n = 2.$

$$F(y) = \begin{cases} \dfrac{10^{3k}}{4} y^2 + 10^{2k} y + 10^k & \text{for } -2 \cdot 10^{-k} < y \leq -10^{-k} \\ -\dfrac{10^{3k}}{4} y^2 + \dfrac{10^k}{2} & \text{for } -10^{-k} \leq y \leq 10^{-k} \\ \dfrac{10^{3k}}{4} y^2 - 10^{2k} y + 10^k & \text{for } 10^{-k} \leq y < 2 \cdot 10^{-k} \\ 0 & \text{elsewhere} \end{cases}$$

$n = 3.$

$$F(y) = \begin{cases} \dfrac{10^{4k}}{54} y^3 + \dfrac{2 \cdot 10^{3k}}{9} y^2 + \dfrac{8 \cdot 10^{2k}}{9} y + \dfrac{32 \cdot 10^k}{27} & \\ \qquad\qquad \text{for } -4 \cdot 10^{-k} < y \leq -3 \cdot 10^{-k} & \\ -\dfrac{10^{4k}}{54} y^3 - \dfrac{10^{3k}}{9} y^2 - \dfrac{10^{2k}}{9} y + \dfrac{5 \cdot 10^k}{27} & \text{for } -3 \cdot 10^{-k} \leq y \leq -2 \cdot 10^{-k} \\ \dfrac{10^{2k}}{9} y + \dfrac{10^k}{3} & \text{for } -2 \cdot 10^{-k} \leq y \leq -10^{-k} \end{cases}$$

$n = 3$.—*cont.*

$$F(y) = \text{(cont.)} \begin{cases} -\frac{10^{4k}}{27}y^3 - \frac{10^{3k}}{9}y^2 + \frac{8\cdot 10^k}{27} & \text{for} \quad -10^{-k} \le y \le 0 \\ \frac{10^{4k}}{27}y^3 - \frac{10^{3k}}{9}y^2 + \frac{8\cdot 10^k}{27} & \text{for} \quad 0 \le y \le 10^{-k} \\ -\frac{10^{2k}}{9}y + \frac{10^k}{3} & \text{for} \quad 10^{-k} \le y \le 2\cdot 10^{-k} \\ \frac{10^{4k}}{54}y^3 - \frac{10^{3k}}{9}y^2 + \frac{10^{2k}}{9}y + \frac{5\cdot 10^k}{27} & \text{for} \quad 2\cdot 10^{-k} \le y \le 3\cdot 10^{-k} \\ -\frac{10^{4k}}{54}y^3 + \frac{2\cdot 10^{3k}}{9}y^2 - \frac{8\cdot 10^{2k}}{9}y + \frac{32\cdot 10^k}{27} & \text{for} \quad 3\cdot 10^{-k} \le y < 4\cdot 10^{-k} \\ 0 & \text{elsewhere.} \end{cases}$$

In general, $F(y)$ is an even continuous function, vanishing for $|y| \ge 2^{n-1}10^{-k}$ and defined by different n^{th} degree polynomials in different intervals.

The above frequency functions were derived on the assumption that the x's are random numbers drawn from a rectangular distribution. However, the results may be applied to the entries of a mathematical table provided the rounding errors are horizontally distributed and the difference under consideration is of such an order that the digits in the decimal place corresponding to the last place given in the table are also horizontally distributed. These conditions are frequently satisfied. Since data on the errors in the second differences of a table of $Ci(x) = \int_\infty^x \frac{\cos x}{x}\,dx$ given to 9 decimals was available, a study was made of a sample of 1000 of these errors. The theoretical and observed frequencies for this sample are given in the following table:

Error	−2 to −1.5	−1.5 to −1.0	−1.0 to −.5	−.5 to 0	0 to .5	.5 to 1.0	1.0 to 1.5	1.5 to 2	
Theoretical Frequency	10.4	72.9	177.1	239.6	239.6	177.1	72.9	10.4	1000.0
Observed Frequency	9	68	161	272	243	174	63	10	1000

By applying Pearson's χ^2-test, it is found that the observed frequencies show no significant deviations from the theoretical frequencies

NEW YORK

ON TESTING THE HYPOTHESIS THAT TWO SAMPLES HAVE BEEN DRAWN FROM A COMMON NORMAL POPULATION

By B. A. Lengyel[1]

1. **Introduction.** This paper is devoted to the problem of testing the hypothesis that two samples of 2, 3 and 4 variables, and of equal size, have been drawn from a common unspecified normal population. It is, in a certain sense, a continuation of J. W. Fertig's papers [1, 2] which were devoted to the problem of testing the hypothesis that one or more samples of n variables have been drawn from a completely or partially specified normal population.

For the sake of application to biological research, it is important to have means of determining whether two samples may have come from a common population even if this population is unknown. Moreover, it is often imperative to test two samples with respect to all their variables simultaneously. Much valuable information may be lost if the variables are tested individually. One has to consider not only the fact that two samples which differ almost significantly from each other in each variable separately might be significantly different if the probabilities would be combined, but one has to take account of the possible correlations between the variables which are completely disregarded if the tests are applied to each variable separately. It is not difficult to imagine two samples of two variables with identical means and variances and significantly different correlation coefficients.

J. Neyman and E. S. Pearson [3] have investigated the problem of testing statistical hypotheses in general. They have developed the method of likelihood ratios. It is beyond the scope of the present paper to give an account of this theory; we have to restrict ourselves to statements concerning the fundamental concepts we are going to apply to our specific problem.

A sample with one variable and of size N can be regarded as a point in an N-dimensional space. The acceptance or rejection of a hypothesis concerning this sample will depend on whether or not the point representing the sample is contained in certain critical regions determined by the hypothesis and by the statistical criterion that is to be applied The choice of the critical regions is of fundamental importance; its implications have been thoroughly discussed by Neyman and Pearson. These authors found a useful criterion for testing the hypothesis that a sample was drawn from a specified member of an admissible set of populations by introducing the ratio of the likelihood that the sample was drawn from the specified population to the maximum value of the likelihood for all populations in the admissible set (Cf. §2). This ratio λ can vary between

[1] From the Research Service of the Worcester State Hospital, Worcester, Massachusetts

0 and 1. The association between values of λ and the credibility of the hypothesis in question is such that the greater the value of λ the greater the degree of tenability of the hypothesis. λ = constant defines a surface in the sample space. These surfaces are the contours of the critical regions associated with the acceptance or rejection of the hypothesis A hypothesis is rejected as untenable if λ is so small that

$$\int_0^{\lambda} P(\lambda)\, d\lambda < \alpha,$$

where α is some value arbitrarily small, say 01 or .05 and $P(\lambda)$ is the distribution of λ if the hypothesis is true.

This method of testing hypotheses is evidently not restricted to one sample with one variable, nor is it restricted to *simple* hypotheses. A simple hypothesis is one which is associated with one completely specified population. A composite hypothesis is one which is associated with a subset of the admissible populations. For example, the hypothesis that a sample with n variables has been drawn from a normal population with means a_1 , a_2 , a_n whatever may be the variances and correlation coefficients is a composite hypothesis. Such is the hypothesis that two or more samples have come from a common but unspecified population.

The problem of several samples with one variable was discussed by Neyman and Pearson [4, 5]; the problem of several samples with two variables by Pearson and Wilks [6]. In another paper Wilks [7] derived formulas for λ and the moments of $P(\lambda)$ for the most general case of k samples of n variables. For the sake of practical applications it is necessary to have tables for the limits of significance of λ. Such tables have been prepared for samples with one variable by Neyman and Pearson and for completely or at least partially specified populations and more than one variable by Fertig [1, 2]. The present paper contains tables for the case of 2, 3 and 4 variables and a common unspecified normal population. Since the case of two variables has been theoretically solved by Pearson and Wilks we shall have to compare our results with those of the above authors who derived the distribution of $P(\lambda)$ but did not compute tables.

Our procedure is the following: We start with the moments of $P(\lambda)$ as given by Wilks and approximate the distribution of $\lambda^{1/N}$ by a suitable function. Then we compute the limits of significance for this approximating function. This procedure was originally suggested by Neyman and Pearson and was applied with some modifications by Fertig.

§2 contains the definition of the likelihood ratio λ; §3 deals with the moments of its distribution for the case of a common unspecified population. In §4 we introduce the approximating distribution $y = Cx^{p-1}(1 - x)^{q-1}$ and discuss the determination of the parameters p and q. In §5 we give an independent derivation of the formula obtained by Pearson and Wilks for $P(\lambda)$ for the case of two samples with two variables and compare our approximation with the exact

formula. §6 deals with the determination of the limits of significance and contains the tables. §7 is devoted to an example.

2. **Definition of λ.** Let C_π denote the probability of obtaining a given sample from a population π. C will depend on the parameters of the population and the data of the sample. Let Ω be the set of all admissible populations and ω a subset of Ω which corresponds to a certain hypothesis that is to be tested. Intuitively one would consider a hypothesis tenable or plausible if it gives a high probability density for the given sample if compared with other possible hypotheses. Following this reasoning Neyman and Pearson defined the likelihood of a hypothesis as the ratio of $\underset{\pi \in \omega}{\text{Max}}\, C_\pi$ to $\underset{\pi \in \Omega}{\text{Max}}\, C_\pi$. In the special case which we propose to investigate, the populations are assumed to be normal. We wish to test the hypothesis that two given samples have come from a *common* unspecified population. Hence λ, the likelihood of this hypothesis, is the maximum likelihood that the samples have come from a common normal population divided by the maximum likelihood that the samples have come from any two normal populations.

The value of λ can be expressed by the variates of the samples by means of the following formula [Cf. [7] p. 489]

$$\lambda = \left[\frac{S_1}{S_0}\right]^{\frac{1}{2}N_1} \left[\frac{S_2}{S_0}\right]^{\frac{1}{2}N_2}, \tag{1}$$

where S_1 and S_2 are the generalized variances[2] of the samples and S_0 is the generalized variance of the sample obtained by pooling the two given samples. N_1 is the size of the first sample, N_2 the size of the second. In case of equal samples to which we shall restrict ourselves $N_1 = N_2 = N$; thus

$$\lambda^{1/N} = \frac{S_1^{\frac{1}{2}} S_2^{\frac{1}{2}}}{S_0}. \tag{2}$$

3. **The Moments of the λ Distribution.** The distribution of λ depends on the number of variables, the number and the size of the samples and on the kind of hypothesis that is to be tested; e.g. that the samples have come from a common unspecified population. This distribution has been evaluated for the case of equal samples of one and two variables and our hypothesis concerning a common unspecified population. The general form of this distribution is still unknown and even the known formula for two variables is not very suitable for computation. Therefore we shall follow the procedure introduced by Neyman and Pearson [4] and we shall use the known moments of the unknown distribution

[2] The generalized variance of a sample is a determinant, the elements of which are the variances and covariances Thus, for two variables x and y the generalized variance $S = S_x^2 S_y^2 (1 - r^2)$; where S_x^2 and S_y^2 denote the variances of x and y respectively, r the correlation coefficient.

function $P(\lambda)$ in order to construct an approximation to $P(\lambda)$. For two equal samples of n variables the moments of $P(\lambda)$ about the origin are [Cf. [7] p. 490]

$$(3) \qquad M_h = 2^{nNh} \prod_{i=1}^{n} \left\{ \frac{\Gamma\left(\frac{N(1+h)-i}{2}\right)}{\Gamma\left(\frac{N-i}{2}\right)} \right\}^2 \frac{\Gamma\left(\frac{2N-i}{2}\right)}{\Gamma\left(\frac{2N(1+h)-i}{2}\right)}$$

$$\text{for } h = 1, 2, 3, \cdots.$$

Equation (2) readily suggests that we should compute or approximate the distribution of $\lambda^{1/N}$ rather than that of λ. Let μ_2 denote the h-th moment of $P(\lambda^{1/N})$ then $\mu_h = M_{h/N}$ follows immediately from the definition of $M_h = \int_0^1 \lambda^h P(\lambda) d\lambda$. Hence in order to obtain the μ's we have to replace Nh by h in (3).

$$(4) \qquad \mu_h = 2^{nh} \prod_{i=1}^{n} \left\{ \left[\frac{\Gamma\left(\frac{N+h-i}{2}\right)}{\Gamma\left(\frac{N-i}{2}\right)} \right]^2 \frac{\Gamma\left(\frac{2N-i}{2}\right)}{\Gamma\left(\frac{2N+2h-i}{2}\right)} \right\}.$$

This expression can be much simplified for all given values of h and n. However, there is no need for such simplification, because one has to compute the first moments only. All higher moments can be expressed by means of the first moments for various N's. The dependence of μ_h on N is evident from (4), we shall indicate it by writing $\mu_h(N)$. The ratio of two subsequent moments is

$$(5) \qquad \frac{\mu_{h+1}(N)}{\mu_h(N)} = 2^n \prod_{i=1}^{n} \left\{ \left[\frac{\Gamma\left(\frac{N+h+1-i}{2}\right)}{\Gamma\left(\frac{N+h-i}{2}\right)} \right]^2 \frac{\Gamma\left(\frac{2(N+h)-i}{2}\right)}{\Gamma\left(\frac{2(N+h+1)-i}{2}\right)} \right\}$$

$$= \mu_1(N+h).$$

Equation (5) contains an important relation of the moments. In fact from (5) follows:

$$(6) \qquad \begin{cases} \mu_2(N) = \mu(N)\mu(N+1), \\ \mu_3(N) = \mu(N)\mu(N+1)\mu(N+2), \\ \vdots \\ \mu_h(N) = \mu(N)\mu(N+1)\cdots\mu(N+h-1), \end{cases}$$

where the 1's from $\mu_1(N)$'s have been omitted. This last group of equations holds for any number of variables. Thus we have to compute $\mu(N)$ for each N, then multiplication gives the higher moments.

For[3] $n = 2$, $$\mu(N) = \frac{(N-2)^2}{(N-1)(N-\frac{1}{2})}$$

For $n = 3$, $$\mu(N) = 2^3\left[\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-3}{2}\right)}\right]^2 \frac{1}{(N-\frac{1}{2})(N-1)(N-\frac{3}{2})}.$$

For $n = 4$, $$\mu(N) = \frac{(N-2)^2(N-4)^2}{(N-1)(N-\frac{1}{2})(N-2)(N-\frac{3}{2})}.$$

4. **Approximation of the distribution of $\lambda^{1/N}$.** Following the procedure of Neyman and Pearson we shall use the moments computed in the previous section for the fitting of a Pearson frequency curve to the unknown distribution $P(\lambda^{1/N})$. Since $0 \leqq \lambda \leqq 1$ it is natural to fit a frequency curve of the following type

$$y = Cx^{p-1}(1-x)^{q-1}, \tag{7}$$

where $$C = \frac{1}{B(p,q)} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}.$$

The first two moments are sufficient to determine the two parameters p and q. The moments of the distribution (7) are readily computed:

$$\nu_1 = \frac{p}{p+q}, \qquad \nu_2 = \nu_1\frac{p+1}{p+q+1} = \frac{p}{p+q}\cdot\frac{p+1}{p+q+1}. \tag{8}$$

In general:

$$\frac{\nu_{h+1}}{\nu_h} = \frac{p+h}{p+q+h}. \tag{9}$$

Equation (9) corresponds to equation (5) since one can write $\nu = \nu(p, q)$, then (9) becomes

$$\frac{\nu_{h+1}(p,q)}{\nu_h(p,q)} = \nu(p+h, q). \tag{10}$$

At first sight the similarity of equations (5) and (10) would suggest that one should choose p and q so that $\nu(p+h, q) = \mu(N+h)$, for $h = 0, 1, 2, 3, \cdots$. However, this cannot be done because the equations which express the equality of the first two moments:

$$\nu(p, q) = \mu(N) \tag{11}$$

$$\nu(p+1, q) = \mu(N+1) \tag{12}$$

[3] The case $n = 1$ is omitted here since it has been treated by Neyman and Pearson [4].

determine p and q *completely*. The quantities $\nu(p + h, q)$ can be only *approximately* equal to $\mu(N + h)$ for $h > 1$. The goodness of fit may be tested by comparing the third and fourth moments.

The advantage of equations (5) and (10) is that once p and q have been computed one does not have to compare

$$\nu_3 = \nu_2 \nu(p + 3, q)$$

with

$$\mu_3 = \mu_2 \mu(N + 3),$$

but since $\nu_2 = \mu_2$ one can compare $\nu(p + 3, q)$ with $\mu(N + 3)$. Similarly the comparison of the fourth moments can be replaced by comparing $\nu(p + 4, q)$ with $\mu(N + 4)$. It has to be remembered that once the sequence of $\mu(N)$'s has been computed for all N's, each of its terms can be used four times in the determination and the checking of p's and q's.

The general procedure for the determination of p and q is to compute the $\mu(N)$'s first and then solve the equations (11) and (12): i.e.

$$\frac{p}{p + q} = \mu(N), \tag{13}$$

$$\frac{p + 1}{p + q + 1} = \mu(N + 1). \tag{14}$$

The solution of these equations is:

$$p = \mu(N) \frac{1 - \mu(N + 1)}{\mu(N + 1) - \mu(N)}, \tag{15}$$

$$q = \left[\frac{1}{\mu(N)} - 1\right] p. \tag{16}$$

As N increases $\mu(N)$ approaches 1 from below; $\mu(N + 1) - \mu(N)$ will be very small. E.g., for $n = 2$ as N varies from 30 to 50 $\mu(N)$ increases from .9164 to .9499. It is easily seen that small errors in μ may produce much larger errors in p and q. For $n = 4$ it was necessary to compute μ to nine decimal places to get p and q to three decimal places accurately. For $n = 2$ equations (13) and (14) become

$$\frac{p}{p + q} = \frac{(N - 2)^2}{(N - 1)(N - \frac{1}{2})}, \tag{17}$$

$$\frac{p + 1}{p + q + 1} = \frac{(N - 1)^2}{N(N + \frac{1}{2})}. \tag{18}$$

These can be solved explicitly

$$p = (N - 2)\left[1 - \frac{3(N - 1)}{5N^2 - 9N + 1}\right], \tag{19}$$

$$q = 2.5 + \frac{4.5}{5N^2 - 9N + 1} \tag{20}$$

The last two equations enable us to compute p and q directly and thus avoid the more laborious computation by means of the $\mu(N)$'s. For $n = 3$ and 4, however, such a short cut was not found. The computation of μ's for $n = 4$ was facilitated by the following relation:

$$_4\mu(N) = {}_2\mu(N)\,\frac{(N-4)^2}{(N-2)(N-\frac{3}{2})}, \tag{21}$$

where the suffices denote the number of variables in the problem to which the μ's refer. Thus the computed values of $_2\mu(N)$ were again used. Eight-place logarithms were used in the computation of $_3\mu(N)$ from the formula at the end of §3.

5 **The distribution of $\lambda^{1/N}$ for two variables.** For two variables it is possible to evaluate the distribution of $\lambda^{1/N}$ or some other suitable power of λ directly from the moments. Pearson and Wilks (Cf. [6] pp. 364-368) derived the distribution of $\lambda^{2/N}$ for this case. Their method was adapted to the treatment of more general problems than ours. It is possible to derive the distribution of $\lambda^{1/N}$ in our special case more directly:

For $n = 2$ the moments of $\lambda^{1/N}$ are:

$$\mu_h = 2^{2h}\left[\frac{\Gamma\left(\frac{N+h-1}{2}\right)\Gamma\left(\frac{N+h-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)}\right]^2 \frac{\Gamma(N-\frac{1}{2})\Gamma(N-1)}{\Gamma(N+h-\frac{1}{2})\Gamma(N+h-1)}, \tag{22}$$

$$h = 1, 2, 3, \cdots.$$

Applying the following transformation formula[4]

$$\Gamma(z)\Gamma(z+\tfrac{1}{2}) = \frac{\sqrt{\pi}}{2^{2z-1}}\Gamma(2z) \tag{23}$$

to

$$z_1 = \tfrac{1}{2}(N+h-2), \qquad z_2 = \tfrac{1}{2}(N-2), \qquad z_3 = N-1 \quad \text{and} \quad z_4 = N+h-1$$

(22) becomes

$$\mu_h = 2^{2h}\left[\frac{\Gamma(N+h-2)}{\Gamma(N-2)}\right]^2 \frac{\Gamma(2N-2)}{\Gamma(2N+2h-2)}. \tag{24}$$

Thus μ_h is of the form $F(N+h)/F(N)$ with

$$F(N) = 2^{2N}\frac{\Gamma(N-2)^2}{\Gamma(2N-2)}.$$

[4] Cf. Whittaker and Watson. Modern Analysis, 4th ed , p. 240.

Our problem is to find a function $P(t)$ such that

$$\mu_h = \int_0^1 t^h P(t)\,dt, \tag{25}$$

$$\text{for } h = 0, 1, 2, \cdots .$$

This problem is solved if we can find a function of N and t, say, $p(N, t)$ such that

$$F(N + h) = \int_0^1 t^h p(N, t)\,dt \tag{26}$$

$$\text{for } h = 0, 1, 2, \cdots .$$

N and h enter the left side of equation (26) symmetrically. The same must be true for the right side. Hence $p(N, t)$ must have the form $t^N f(t)$ where $f(t)$ is independent of N. If then (26) is satisfied for all N and $h = 0$, it is also satisfied for all N and all h.

Let us now examine $F(N)$. Applying again the transformation (23) we can bring it to the form:

$$\begin{aligned} F(N) &= 2^3 \frac{\Gamma(N-2)\sqrt{\pi}}{\Gamma(N-\tfrac{1}{2})(N-2)} = 2^4 \frac{\Gamma(N-2)\Gamma(\tfrac{3}{2})}{(N-2)\Gamma(N-\tfrac{1}{2})} \\ &= \frac{2^4}{N-2} B(N-2, \tfrac{3}{2}). \end{aligned} \tag{27}$$

Now $B(N-2, \frac{3}{2})$ can be represented as an integral of the desired type

$$B(N-2, \tfrac{3}{2}) = \int_0^1 t^{N-3}(1-t)^{\frac{1}{2}}\,dt. \tag{28}$$

We set $p(N, t) = 2^4 t^{N-3} g(t)$ and seek to determine $g(t)$ so that (26) will be satisfied for all N and for $h = 0$: i.e.,

$$F(N) = \frac{2^4}{N-2} B(N-2, \tfrac{3}{2}) = 2^4 \int_0^1 t^{N-3} g(t)\,dt. \tag{29}$$

An integration by parts with $g(1) = 0$ gives

$$\frac{2^4}{N-2} B(N-2, \tfrac{3}{2}) = -\frac{2^4}{N-2} \int_0^1 t^{N-2} g'(t)\,dt \tag{30}$$

This equation evidently holds for all N if and only if

$$-tg'(t) = \sqrt{1-t}. \tag{31}$$

This differential equation is readily solved by the substitution of $y = 1 - t$. In fact it becomes

$$\frac{dg(y)}{dy} = \frac{2y^2}{1-y^2} = 2[y^2 + y^4 + y^6 + \cdots]. \tag{32}$$

Hence

$$g = \log \frac{1+y}{1-y} - 2y = 2\left[\log \frac{1+\sqrt{1-t^2}}{\sqrt{t}} - \sqrt{1-t^2}\right]. \tag{33}$$

The complete solution for the distribution function is

$$dP(t) = \frac{\Gamma(2N-2)}{2^{2N-4}[\Gamma(N-2)]^2}\, t^{N-3}\left[\log \frac{1+\sqrt{1-t^2}}{\sqrt{t}} - \sqrt{1-t^2}\right] dt \tag{34}$$

with $t = \lambda^{1/N}$ in accordance with the formula of Pearson and Wilks.[5] Integration by parts gives

$$P(\lambda^{1/N} < t) = \frac{\Gamma(2N-2)}{2^{2N-5}\Gamma(N-1)\Gamma(N-2)} \cdot \left\{t^{N-2}\left[\log \frac{1+\sqrt{1-t^2}}{\sqrt{t}} - \sqrt{1-t^2} + \tfrac{1}{2}\int_0^t y^{N-3}\sqrt{1-y}\, dy\right\}. \tag{35}$$

One can use this last equation to determine the limits of significance. However, this was not done when the tables of this paper were computed. The approximation of the distribution function by the function described in equation (7) was deemed sufficient and the use of the tables of the incomplete beta function greatly facilitated the computation.

In concluding this section we wish to demonstrate the goodness of approximation of the exact distribution function by a function of the type $Ct^{p-1}(1-t)^{q-1}$ with p and q given by equations (19) and (20).

For small values of t the shape of the curve is determined by the exponent of t, which is exactly $N-3$ for the distribution function and nearly $N - 3 - \frac{3}{8}$ for the approximating function. For large t; i.e., small $1-t$, the exponent of $(1-t)$ is the determining factor. By (32) we have

$$g(\sqrt{1-t}) = 2\left[\frac{(1-t)^{\frac{1}{2}}}{3} + \frac{(1-t)^{\frac{1}{2}}}{5} + \cdots\right],$$

or approximately $\frac{4}{3}(1-t)^{\frac{1}{2}}$. For the approximating curve $q - 1 = \frac{3}{2} + O(1/N^2)$ which is even better agreement. It is easily seen that the goodness of approximation increases with N.

6. **Determination of the Levels of Significance.** The final task was to compute the values of x which satisfy the equations:

$$I_x(p, q) = \frac{1}{B(p, q)}\int_0^x t^{p-1}(1-t)^{q-1}\, dt = \alpha,$$

with $\alpha = .01$ and $\alpha = 05$. This was done by interpolation in the *Tables of the Incomplete Beta Function* [8]. In these tables the argument x increases by steps of .01. A value x_0 was determined by inspection, so that $I_{x_0}(p, q) < \alpha$ but $I_{x_1}(p, q) > \alpha$ where $x_1 = x_0 + .01$. The values of $I_{x_0}(p, q)$ and $I_{x_1}(p, q)$ were

[5] Cf. [6] p. 368 Equation 60 ($l^2 = t$).

determined by interpolation with respect to p and q, using the two-dimensional Everett formula, neglecting fourth and higher differences. x was then determined by linear interpolation. It is worth while to mention that the terms of second order in Everett's formula decreased quite rapidly as N increased. Once this was noticed some labor has been saved by not computing the terms of second order for values of N between 30 and 50 but by estimating the second order terms from those obtained for $N = 30, 40$ and 50.

Levels of Significance of $\lambda^{1/N}$

Sample Size N	2 Variables		3 Variables		4 Variables	
	1%	5%	1%	5%	1%	5%
10	.395	.507	.238	.328	.122	.184
11	.437	.546	.282	.374	.153	.217
12	.475	.579	.323	.414	.198	.270
13	.508	.610	.360	.451	.233	.308
14	.537	.634	.393	.483	.267	.343
15	.563	.656	.423	.512	.298	.375
16	.586	.676	.451	.537	.328	.405
17	.607	.694	.476	.561	.355	.432
18	.626	.710	.500	.582	.380	.456
19	.644	.724	.521	.601	.404	.479
20	.660	.737	.541	.619	.426	.501
22	.687	.760	.576	.650	.466	.538
24	.711	.779	.606	.676	.501	.571
26	.731	.795	.632	.699	.532	.599
28	.749	.809	.655	.719	.560	.624
30	.765	.821	.675	.736	.584	.646
32	.778	.832	.694	.752	.606	.666
34	.791	842	.710	.765	.626	.683
36	.802	.850	.724	.778	.644	.700
38	.811	.858	.738	.789	.660	.714
40	.820	.865	.750	.799	.676	.727
42	.828	.871	.761	.808	.689	.739
44	.836	.877	.771	.816	.702	.750
46	.843	.882	.780	.824	.713	.760
48	.849	.887	.789	.831	.724	.769
50	.854	.891	.796	.837	.734	.773

7. **An Example.** The problem chosen to illustrate the use of the tables is taken from a study on insulin-treated schizophrenic patients of the Worcester State Hospital. It was attempted to differentiate between those patients who recovered after treatment and those who did not recover. Blood constituents and blood pressure were determined among other variables.

The variables in this example are designated as x = blood phosphorus, y = cholesterol in mg./100 cc., z = blood pressure in mm. Hg. The statistics for the 10 "recovered" patients are:

$$S_x^2 = 2.222 \qquad S_y^2 = 376.50 \qquad S_z^2 = 51.97$$

$$r_{12}S_xS_y = -1.121 \qquad r_{13}S_xS_z = -8.217 \qquad r_{23}S_yS_z = 12.51$$

For "not-recovered" 10 patients

$$S_x^2 = 3.120 \qquad S_y^2 = 816.19 \qquad S_z^2 = 96.32$$

$$r_{12}S_xS_y = 26.23 \qquad r_{13}S_xS_z = 2.92 \qquad r_{23}S_yS_z = 65.78$$

For the total group of 20

$$S_x^2 = 3.034 \qquad S_y^2 = 609.02 \qquad S_z^2 = 83.09$$

$$r_{12}S_xS_y = 10.41 \qquad r_{13}S_xS_z = -.845 \qquad r_{23}S_yS_z = 15.99$$

These values give for the sample variances 17,462; 168,628; and 143,904, respectively.

Hence

$$\lambda^{1/10} = \frac{\sqrt{17{,}462 \times 168{,}628}}{143{,}904} = .377.$$

The 5% limit of significance is .328, hence the two groups do not differ significantly from each other.

REFERENCES

[1] J. W. Fertig, "Tests of Significance for Multivariate Samples," *Annals of Math. Stat.* Vol. 7 (1936) pp. 113–121.

[2] J. W. Fertig, "Testing of Certain Hypotheses by Means of Lambda Criteria," *Biometric Bulletin* Vol 1, pp. 45–82.

[3] J. Neyman and E. S Pearson, "Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika* 20A (1928) pp. 175–241.

[4] J. Neyman and E. S. Pearson, "The Problem of Two Samples," *Bull. Int. Acad. Polonaise* Series A (1931) pp. 73–96.

[5] J. Neyman and E. S. Pearson, "The Problem of k Samples," *Bull. Int. Acad. Polonaise* Series A (1931) pp. 260–281.

[6] E S. Pearson and S. S. Wilks, "Methods of Statistical Analysis Appropriate for k Samples of Two Variables," *Biometrika* Vol 25 (1933) pp. 353–378.

[7] S. S. Wilks, "Certain Generalizations of the Analysis of Variance," *Biometrika* Vol. 24 (1932) pp. 471–494.

[8] *Tables of the Incomplete Beta Function*, London, (1934).

Rensselaer Polytechnic Institute,
Troy, N. Y

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

THE DISTRIBUTION OF "STUDENT'S" RATIO FOR SAMPLES OF TWO ITEMS DRAWN FROM NON-NORMAL UNIVERSES

By Jack Laderman

The fundamental assumption in the derivation of "Student's" distribution[1]

$$f(t)\,dt = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)\left(1+\frac{t^2}{n-1}\right)^{\frac{1}{2}n}}\,dt$$

is that the universe sampled is normal. When the universe sampled is non-normal and n is small, the distribution of t differs considerably from "student's" distribution. In 1929, Rider[2] derived the distribution of t for samples of two drawn from the rectangular distribution

$$f(x)\,dx = \begin{cases} 1\,dx & \text{for } 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

In this paper, the formal expression for the distribution of t will be derived for samples of two drawn from any population having a continuous frequency function. A geometrical method similar to that employed by Rider will be used

Let the universe sampled have the frequency function, $f(x)$, with zero mean, and let $f(x)$ be greater than zero from $x = a$ to $x = b$ and equal to zero elsewhere. Suppose the two observations are x_1 and x_2.

Then $$\bar{x} = \frac{x_1 + x_2}{2}$$

and $$t = \frac{\sqrt{n}\,\bar{x}}{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}} = \frac{\sqrt{2}\,\bar{x}}{\sqrt{(x_1-\bar{x})^2 + (x_2-\bar{x})^2}}.$$

[1] "Student", "The Probable Error of a Mean" *Biometrika*, Vol. VI (1908), pp. 1-25

[2] Paul R Rider, "On the Distribution of the Ratio of the Mean to Standard Deviation in Small Samples from Non-Normal Universes", *Biometrika*, Vol. XXI (1929), pp. 124-143.

The sample (x_1, x_2) can be represented by a point in a square of side $b-a$, as point P in Figure 1.

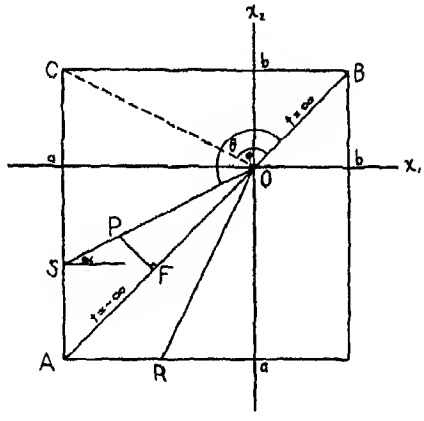


FIG. 1

The coordinates of F are $(\bar{x}, \bar{x})$

$$OF = -\sqrt{2}\,\bar{x}$$

$$FP = \sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2}$$

therefore
$$t = -\frac{OF}{FP} = \cot\theta.$$

Similarly for a point lying below AB, the value of t is $-\cot\theta$. Hence all points on OS and its image OR have the same value of t.

Let
$$\alpha = \theta - \frac{3\pi}{4}$$

Then
$$\tan\alpha = \frac{t+1}{t-1}$$ and the equation of OS is

$$x_2 = \frac{t+1}{t-1}\,x_1.$$

The probability of getting a sample point in the element of area $dx_1\,dx_2$ is $f(x_1)f(x_2)\,dx_1\,dx_2$. Therefore the probability of getting a value of t less than the value represented by a point on OS is given by

$$(1) \qquad 2\int_a^0 \int_{x_1}^{\frac{t+1}{t-1}x_1} f(x_1)f(x_2)\,dx_2\,dx_1.$$

By differentiating (1), we get the frequency function

$$y(t) = -\frac{4}{(t-1)^2}\int_a^0 x_1 f(x_1) f\left(\frac{t+1}{t-1}\,x_1\right) dx_1.$$

However, this expression is valid only when $t \leq \cot \varphi$ where φ is the angle between OB and OC in Figure 1. From Figure 1 we notice that

$$\varphi = \frac{\pi}{4} + \cot^{-1}\left(-\frac{b}{a}\right)$$

$$\text{therefore } \cot \varphi = \frac{b+a}{b-a}.$$

Hence the above expression, $g(t)$, is valid for $t \leq \frac{b+a}{b-a}$.

When $t \geq \frac{b+a}{b-a}$, the probability of obtaining a value of t greater than the

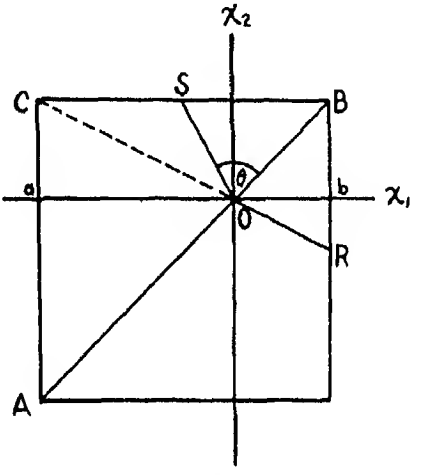


FIG. 2

value represented by a point on OS or its image OR, as in Figure 2, is given by

$$2\int_0^b \int_{\frac{t-1}{t+1}x_2}^{x_2} f(x_1)f(x_2)\,dx_1\,dx_2$$

and the distribution function is

$$1 - 2\int_0^b \int_{\frac{t-1}{t+1}x_2}^{x_2} f(x_1)f(x_2)\,dx_1\,dx_2. \tag{2}$$

After differentiating (2), we obtain the frequency function

$$g(t) = \frac{4}{(t+1)^2}\int_0^b x_2 f(x_2) f\left(\frac{t-1}{t+1}x_2\right) dx_2.$$

Thus, the frequency function of t for samples of two can be obtained from the expressions:

$$(3) \qquad g(t) = \begin{cases} \dfrac{4}{(t-1)^2}\displaystyle\int_0^a xf(x)f\left(\frac{t+1}{t-1}x\right)dx & \text{for } t \leq \dfrac{b+a}{b-a} \\ \dfrac{4}{(t+1)^2}\displaystyle\int_0^b xf(x)f\left(\frac{t-1}{t+1}x\right)dx & \text{for } t \geq \dfrac{b+a}{b-a} \end{cases}$$

These expressions may also be used when a, or b, or both are infinite. However, the join point $\dfrac{b+a}{b-a}$ is then indeterminate, but by consideration of Figure 1, it can easily be seen that the join points are as follows:

$a = -\infty$ b finite	$t = -1$
a finite $b = +\infty$	$t = 1$
$a = -\infty$ $b = +\infty$	$t = 0$

The expressions given by (3) have been verified by obtaining the distribution of t for samples of two drawn from the normal distribution and also from the rectangular distribution. The explicit distributions were found very easily from (3) by performing the integrations.

For instance when $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$$

we get $$g(t) = \frac{1}{\pi(1+t^2)} \qquad -\infty < t < +\infty$$

which agrees with Student's distribution for $n = 2$.

And when $$f(x) = \begin{cases} 1 & -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

we get $$g(t) = \begin{cases} \dfrac{1}{2(t-1)^2} & \text{for } t \leq 0 \\ \dfrac{1}{2(t+1)^2} & \text{for } t \geq 0 \end{cases}$$

which agrees with the distribution found by Rider as corrected by Perlo.[3]

NEW YORK

[3] Victor Perlo, "On the Distribution of Student's Ratio for Samples of Three Drawn from a Rectangular Distribution," *Biometrika*, Vol. XXV (1933) pp. 203–204.

ON SOME INFINITE SERIES INTRODUCED BY TSCHUPROW

By J. B. D. Derksen

In his fundamental work on the principles of the theory of correlation Tschuprow introduces some infinite series, leaving certain questions regarding their convergence or divergence unsolved.[1]

As will be shown in the following note, these series are what may be termed randomly divergent,[2] that is series involving random variables which may take on values which will make the series divergent. This result is of importance: e.g. the well-known formula for the standard error of a correlation coefficient $\left(\frac{1-r^2}{\sqrt{N}}\right)$ is the first term of an infinite series for which the question of convergence has not been carefully considered.

Tschuprow finds himself confronted by infinite series, when dealing with the mathematical expectations of quotients as e.g. correlation coefficients or sums of quotients as e.g. the mean square contingency. Let us consider a two-dimensional discontinuous universe, where the variables are x and y. Let p_{ij} be the probability of the occurrence of the pair of values x_i, y_j. The probability that x assumes the value x_i equals $\sum_{j=1}^{l} p_{ij} = p_{i|}$ $(i = 1, \cdots k; j = 1, \cdots l)$. When taking a sample of N pairs of observations (x, y) the relative frequency of x_i will be $p'_{i|}$, and that of the pair (x_i, y_j) will be p'_{ij}. In accordance with Tschuprow we put $p'_{ij} = p_{ij} + dp_{ij}$ and $p'_{i|} = p_{i|} + dp_{i|}$, where dp_{ij} and $dp_{i|}$ are random variables.

As one of the simplest cases we consider the mathematical expectation of

$$\left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \frac{p_{ij}^2}{p_{i|}^2}\left(1 + \frac{dp_{ij}}{p_{ij}}\right)^2 \cdot \left(1 + \frac{dp_{i|}}{p_{i|}}\right)^{-2}.$$

Now Tschuprow develops the last factor in an infinite binomial series, getting

$$\left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \frac{p_{ij}^2}{p_{i|}^2}\left\{1 + 2\frac{dp_{ij}}{p_{ij}} + \left(\frac{dp_{ij}}{p_{ij}}\right)^2\right\} \cdot \left\{1 - 2\frac{dp_{i|}}{p_{i|}} + 3\frac{dp_{i|}^2}{p_{i|}^2} \cdots\right\}$$

$$= \frac{p_{ij}^2}{p_{i|}^2}\left[1 + 2\frac{dp_{ij}}{p_{ij}} - 2\frac{dp_{i|}}{p_{i|}} + \frac{dp_{ij}^2}{p_{ij}^2} - 4\frac{dp_{i|} \cdot dp_{ij}}{p_{i|} \cdot p_{ij}} + 3\frac{dp_{i|}^2}{p_{i|}^2} \cdots\right]. \tag{1}$$

He has given general formulae (*Biometrika*, vol. XII, p. 194 (1919)) from which the mathematical expectations of the terms of this infinite series may immediately be found. We get an infinite series containing ascending powers of N

[1] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*, Leipzig-Berlin, 1925, p. 85-97. An English translation was prepared by M. Kantorowitsch (*Principles of the Theory of Correlation*, W Hodge & Co., 1939).

[2] Cf my *Inleiding tot de correlatierekening*, Delft, 1935, p. 88-90.

in the denominator. Finally the convergence or divergence of this series has to be investigated, the problem left unsolved by Tschuprow.

The series expansion of $\left(1 + \frac{dp_{s|}}{p_{s|}}\right)^{-2}$ however, diverges for values $dp_{s|}$ such that $\left|\frac{dp_{s|}}{p_{s|}}\right| \geq 1$ and converges only if

$$\left|\frac{dp_{s|}}{p_{s|}}\right| < 1. \tag{2}$$

This result is not affected by the procedure of the determination of mathematical expectations. For if $f(p'_{i|}, p'_{s,})$ is the probability distribution of $(p'_{i|}, p'_{i,})$, then we have to multiply (1) by this function and to sum for all possible values of $p'_{i|}, p'_{s,}$. As the expressions

$$f(p'_{s|}, p'_{s,}) \frac{p^2_{i,}}{p^2_{i|}} \left(1 + \frac{dp_{i,}}{p_{s,}}\right)^{-2}$$

are always positive, the infinite series, which results from replacing the terms of (1) by their mathematical expectations, will also be divergent.

The same argument is true, when we consider for instance the mathematical expectation of the Pearson-Bravais correlation coefficient. Denoting by μ_{11}, μ_{20}, μ_{02} the population values of the product moment and the second order moment of x and y, and by μ'_{11}, μ'_{20}, μ'_{02} the values observed in a sample, the mathematical expectation of the correlation coefficient may be found from

$$E(r) = E\left[\frac{\mu'_{11}}{(\mu'_{20}\,\mu'_{02})^{\frac{1}{2}}}\right] = E\left[\frac{\mu_{11} + d\mu_{11}}{(\mu_{20} + d\mu_{20})^{\frac{1}{2}}(\mu_{02} + d\mu_{02})^{\frac{1}{2}}}\right]$$

$$= \frac{\mu_{11}}{(\mu_{20}\,\mu_{02})^{\frac{1}{2}}} E\left[\frac{1 + \frac{d\mu_{11}}{\mu_{11}}}{\left(1 + \frac{d\mu_{20}}{\mu_{20}}\right)^{\frac{1}{2}}\left(1 + \frac{d\mu_{02}}{\mu_{02}}\right)^{\frac{1}{2}}}\right]$$

where $d\mu_{11}$, $d\mu_{20}$, and $d\mu_{02}$ are random variables. Tschuprow expands the denominator in binomial series. However if $d\mu_{20}$ and $d\mu_{02}$ take on values such that $\left|\frac{d\mu_{20}}{\mu_{20}}\right| \geq 1$ or $\left|\frac{d\mu_{02}}{\mu_{02}}\right| \geq 1$, these series will again be divergent. Analogous difficulties arise in all other cases, where Tschuprow makes use of binomial expansions.

It should also be remarked that the well-known formulae, given by the Biometric School for the standard errors of regression and correlation coefficients are equal to the mathematical expectation of the first terms of infinite series, which, as explained above, are divergent for certain values of the random variables. Therefore the question arises as to what effect the divergence for some of the values of the random variables has on those formulae.

This question can be cleared up by the introduction of Slutsky's *conditionally aleatory* variables.[3] These are defined as follows. Suppose that an aleatory variable z can assume the values $z_1, z_2, \cdots z_n$, with probabilities $p_1, p_2, \cdots p_n$. Now we put some of these probabilities equal to 0, dividing the remaining ones by $1 - Q$, if Q represents the total of all the reduced probabilities. The variable z then becomes the so-called conditionally aleatory variable z'. Moreover we assume that z converges stochastically to some limit. Then Slutsky has shown, that if Q converges to 0, z' will converge to the same stochastical limit as z. Moreover the ratio of corresponding moments of the distributions of z and z' will tend to unity.

Now let us consider for example

$$z = \left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \left(\frac{p_{ij} + dp_{ij}}{p_{i|} + dp_{i|}}\right)^2.$$

Omitting the values for which $| dp_{i|} | \geq p_{i|}$, we get a conditionally aleatory variable z' instead of z. However, according to the theorem mentioned before z' and z will converge to the same stochastical limit, since the probability that $| dp_{i|} | \geq p_{i|}$ converges stochastically to zero as the number of observations increases indefinitely.

In the same way we consider

$$r = \frac{\mu'_{11}}{(\mu'_{20}\mu'_{02})^{\frac{1}{2}}} = \frac{\mu_{11} + d\mu_{11}}{(\mu_{20} + d\mu_{20})^{\frac{1}{2}}(\mu_{02} + d\mu_{02})^{\frac{1}{2}}}$$

and omit those values for which $| d\mu_{20} | \geq \mu_{20}$ and $| d\mu_{02} | \geq \mu_{02}$.

If now we consider the binomial expansions for the conditionally aleatory variables and determine the mathematical expectations of the terms, these new series will converge. All terms in these convergent series will be smaller than the corresponding terms in Tschuprow's series, because we have omitted the larger values of the dp's and the $d\mu$'s. However if the number of observations increases indefinitely the ratios between corresponding terms tend to unity, because the probabilities, that e.g. $| d\mu_{20} | \geq \mu_{20}$ or $| d\mu_{02} | \geq \mu_{02}$ converge to zero.

Let us now turn again to the infinite series given by Tschuprow (loc. cit. p. 90) for the square of the standard error of a correlation coefficient.

$$\sigma_r^2 = E(r - E(r))^2 = \frac{t_1}{N} + \frac{t_2}{N^2} + \frac{t_3}{N^3} + \cdots \tag{3}$$

Here $t_1, t_2, t_3 \cdots$ represent rather lengthy expressions, for which we may refer to Tschuprow's book (loc. cit. p. 88–90). As we have seen before, this series is randomly divergent. However, by introducing a conditionally aleatory variable in the way described above, expanding it into an infinite series and

[3] E. Slutsky, "Uber stochastische Asymptoten und Grenzwerte," *Metron* 1925, Vol. V. No. 3, p. 79. Also my *Inleiding tot de correlatierekening*, Ch V and VI

determining the mathematical expectations of its terms, we get a convergent series, say:

$$\sigma_r'^2 = \frac{t_1'}{N} + \frac{t_2'}{N^2} + \frac{t_3'}{N^3} + \cdots \tag{4}$$

From Slutsky's theorem, mentioned before, it follows that if N increases the ratio of σ_r^2 and $\sigma_r'^2$ will tend to unity. Moreover, if we take N sufficiently large, it will always be possible to fulfill the following inequalities:

$$\left|\frac{t_k'}{t_k}\right| > 1 - \epsilon_k \qquad (k = 1, 2, \cdots n)$$

where ϵ_k $(k = 1, 2, \cdots n)$ and n are arbitrary. Therefore, when n and N are sufficiently large the ratio between the first n terms of the infinite series (3) and the true value of σ_r^2 will differ from 1 by an arbitrary small number. Though the series (3) is divergent for any N, however large, the first n terms of this series will give an approximation of σ_r^2 by taking N sufficiently large.

In this paper we have shown that the procedures which have been followed by the Biometric School and Tschuprow to establish formulas for the standard errors of correlation and regression coefficients and in analogous problems can be made rigorous by the use of conditionally aleatory variables. It was found that their infinite expansions are divergent for some of the values of the random variables involved, however large the number of observations (N) may be. Yet it could be demonstrated, that the first n terms of these series will give an approximation, as close as is wanted, if N is sufficiently large. For practical purposes the case $n = 1$ is the most important.

NETHERLANDS CENTRAL BUREAU OF STATISTICS,
THE HAGUE

A NOTE ON FIDUCIAL INFERENCE

BY R. A. FISHER

In a recent paper [1] Bartlett has written a further justification of his criticism of the test of significance for the difference between means of two samples from normal populations not supposedly of equal or related variance. This test was originally put forward by W. V. Behrens [2], and later [3] found to be very simply derivable by the method of fiducial probability.

It is unfortunate that Bartlett did not restate his own views on this topic without making misleading allusions to mine. Thus, on p. 135 in [1]:

"It is sufficient to note that the distribution certainly provides us with an exact inference of fiducial type, as Fisher himself confirmed [9], p. 375."

I do now know, and Bartlett does not specify, what unguarded statement of mine could be used to justify this assertion. From the time I first introduced

the word, I have used the term *fiducial probability* rather strictly, in accordance with the basic ideas of the theory of estimation. Several other writers have preferred to use it in a wider application, without the restrictions which I think are appropriate. To all, I imagine, it implies at least a valid test of significance expressible in terms of an unknown parameter, and capable of distinguishing, therefore, those values for which the test is significant, from those for which it is not.

Shortly after Bartlett's alternative approach to the problem was put forward [4], I expressed [5] the following opinion. As this occurs prominently in the summary, indeed on the very page to which Bartlett refers in his quotation above, I cannot suppose he has overlooked it, though evidently he must have missed its meaning. I wrote as follows [5] p. 375:

"The criticism of Behrens' test of significance, recently put forward by Bartlett, on the ground that it differs from a possible alternative test, overlooks the inconsistency of assuming for the unknown variances both (a) fiducial distributions in accordance with the samples observed, and (b) values fixed from sample to sample.

The alternative test of significance proposed involves, when the variance ratio of the two populations sampled is unknown, the choice by lot between the value T, used in Behrens' test, and a second value T', which reverses the order of significance of different possible sets of observations. High values of T' are not, therefore, by themselves evidence of inequality of the means"

I submit that the second paragraph quoted above shows, without further argument, that I rejected Bartlett's proposed test of significance, and therefore that I did not confirm his opinion that it provided "an exact inference of fiducial type." Whether my reasons for doing so were strong or weak is, of course, another matter.

What may have led Bartlett to adopt his test of significance is its formal similarity to one appropriate to a different problem. In 1908 "Student" in his now celebrated paper on "The probable error of a mean" [6] applied his solution to what are known as paired observations. Two treatments A and B are applied each to one of a number of pairs of plots, or other experimental units, the members of each pair being chosen to be in other respects closely comparable, although the circumstances of the different pairs are not necessarily closely alike. In order to allow for any, possibly large, variations in the conditions prevailing in the different pairs, attention is confined to the difference, having regard to sign, supplied by each pair.

Thus, if pairs of measurements $a_1, b_1, a_2, b_2, \cdots$ are obtained, we may write

$$d_k = a_k - b_k,$$

and test the hypothesis that the differences d are a normal sample having zero mean. This hypothesis will be true if a_k and b_k are distributed, by experimental error, in normal distributions having the same mean, even though this mean is not the same for different pairs. It will be true if the variances of a and b from the hypothetical mean of the pair are unequal, provided these variances are the same

from pair to pair. These are the reasons which make the hypothesis that d is normally distributed about zero appropriate for testing the differential effect of the treatments.

If only two pairs are used, "Student's" test reduces to

$$t = \frac{d_1 + d_2}{d_1 - d_2}.$$

There is one degree of freedom, so that t is distributed in Cauchy's distribution

$$df = \frac{1}{\pi} \frac{dt}{1 + t^2}.$$

If, now, the symbols have a different meaning, so that a_1 and a_2 are a sample of two from a single normal distribution, and b_1 and b_2 a second sample from a different population, having by hypothesis an independent variance, Behrens' problem (limited for comparison with Bartlett to samples of 2) is to test whether the two populations can be regarded as having the same mean, or whether there is reason to regard the means also as being different. Note that the pairs 1 and 2 are not supposed to differ in treatment or situation. The difference $a_1 - a_2$ is not to be ascribed partly to differences between the hypothetical means of these pairs, but wholly to the error variance of the observations a, about which it is the only source of information; the like is true of the difference $b_1 - b_2$. The sign of these two differences is arbitrary, only their positive values concern our problem. There is no real correspondence between the suffices assigned to the two pairs of letters. They could be interchanged for a, and not for b, without affecting the problem.

Behrens' test reduces for this case to taking

$$T = \frac{a_1 + a_2 - b_1 - b_2}{|a_1 - a_2| + |b_1 - b_2|},$$

using for the probability function, "Student's" distribution for one degree of freedom. Bartlett's test involves choosing at random between T and T', where

$$T' = \frac{a_1 + a_2 - b_1 - b_2}{||a_1 - a_2| - |b_1 - b_2||}.$$

It will be noticed that, if $|b_1 - b_2| < |a_1 - a_2|$, and if, keeping $b_1 + b_2$ constant, $|b_1 - b_2|$ is *increased*, a change which must make us suspect larger errors, and therefore a lower significance, the value of the difference $|a_1 - a_2| - |b_1 - b_2|$ is continuously diminished, and that of T' continuously increased, without limit. In fact, the probability of exceeding any limit of significance, however high, may be made to exceed 50% by this process. The order of significance of such a series of possible observations is thus reversed. The fact that choosing at random T and T' will give us a quantity which, on the null hypothesis, is distributed in "Student's" distribution is, thus, insufficient to justify its use as a test of significance.

It is also irrelevant, and this may be at the present time the most important point to make, that the sampling distribution of T above is not given by "Student's" distribution, if the populations to which statements of probability refer is supposed to consist of samples taken repeatedly from populations having a fixed variance ratio. Such a supposition, as I noted in the passage quoted above, is inconsistent with the fiducial distributions derived from the samples. Bartlett comes near to discussing this point on p. 136 in [1]. He says:

"While Fisher suggests that this in no way invalidates his fiducial argument, in my view if an inference is to be independent of an unknown parameter, it should in particular be independent of it if we imagine that we are being supplied with pairs of samples, for all of which the ratio has the same value."

In its natural meaning this statement seems to be true. The problem concerns what inferences are legitimate from a unique pair of samples, which supply the data, in the light of the suppositions we entertain about their origin; the legitimacy of such inferences cannot be affected by any supposition as to the origin of other samples which do not appear in the data. Such a population of samples is really extraneous to the discussion. Nor has Bartlett shown that Behrens' inference from a unique pair of samples is so affected. What he seems to rely on is that an aggregate of samples fulfilling the null hypothesis, but drawn from pairs of populations having a fixed variance ratio, will show differences between their means exceeding the limits fixed by the test for significance, with a frequency other than that indicated by the test. This, however, is a circumstance common to all the well known tests of significance, and has been obvious from their very origin.

In "Student's" test for significance, for example, if a sample of n' observations are taken from a population normally distributed about zero, we calculate

$$\bar{x} = \frac{1}{n'} S(x), \qquad n = n' - 1, \qquad s^2 = \frac{1}{n} S(x - \bar{x})^2$$

and count $\bar{x}$ as significant, if

$$\bar{x} > s t_n / \sqrt{n'}$$

where t_n is "Student's" test for n degrees of freedom, corresponding to the level of significance chosen.

However, in repeated samples of n' from a population having a given variance σ^2, it is highly improbable that $\bar{x}$ will exceed the limit assigned with the frequency chosen. The limit it will exceed with this frequency is

$$\sigma t_\infty / \sqrt{n'}$$

which will usually differ from that assigned from the sample. This, however, has not hitherto been considered an adequate reason for calling the test inaccurate, or biased. It is merely a recognition of the fact that, if we did know σ, we could make a better test. Just as, in Behrens' problem, if we knew the relative weights x of the observations in the two samples, we could make a

weighted "Student's" test, and should be wise to do so—if the information were available.

Naturally, it may be said that although the limit of significance assigned to $\bar{x}$ will not be verified in repeated sampling from populations having the same variance, the distribution of t will be so verified. In this respect the distribution of t in "Student's" case is analogous to the simultaneous distribution of t_1 and t_2 in Behrens' case, where

$$t_1 = \frac{\bar{x}_1 - \mu}{s_1}, \qquad t_2 = \frac{\bar{x}_1 - \mu}{s_2},$$

μ is the hypothetical common mean of the two populations, and s_1^2, and s_2^2 are the estimated variances of the means of the two samples. The quantity d which Sukhatmé [7] has conveniently tabulated, in such a way that

$$d\sqrt{s_1^2 + s_2^2}$$

supplies a significance limit for $\bar{x}_1 - \bar{x}_2$, naturally does not possess the property that

$$\bar{x}_1 - \bar{x}_2 > d\sqrt{s_1^2 + s_2^2}$$

with the probability assigned, in a population consisting of pairs of samples from populations having the same variance ratio.

If the populations were fixed, the corresponding limit would be

$$t_\infty \sqrt{\sigma_1^2 + \sigma_2^2},$$

and if the variance ratio were fixed so that w is the weight of $\bar{x}_2$ relative to that of $\bar{x}_1$, it would be

$$t_{n_1+n_2} \sqrt{\frac{(n_1 s_1^2 + w n_2 s_2^2)\left(1 + \frac{1}{w}\right)}{n_1 + n_2}}$$

provided always, if we wish to express ourselves in terms of repeated sampling, that the absolute values of σ_1, or σ_1 were fiducially distributed. Behrens' problem refers to the case in which neither the variances nor their ratio is known, so that the unknown variances, independently, must be given their fiducial distributions.

In this note I have not touched on the logical background of Behrens' test, or the practical conditions on which it is appropriate, since I have recently discussed these more fully [8]. Recently also [9] Yates has given a careful explanation of the basis of the test.

SUMMARY

The statement of Bartlett that the author (Fisher) has confirmed that Bartlett's approach to Behrens' problem provides an exact inference of fiducial type is incorrect. The only exact test appropriate to his problem seems to be that given by Behrens.

REFERENCES

[1] M. S. Bartlett, "Complete simultaneous fiducial distributions," *Annals of Math. Stat.*, Vol. X(1939), pp. 129–138.

[2] W.-V Behrens, "Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen," *Landw. Jb.*, Vol. LXVIII(1929), pp. 807–37

[3] R. A. Fisher, "The fiducial argument in statistical inference," *Ann. Eugen*, Vol. VI(1935), pp. 91–98.

[4] M. S. Bartlett, The information available in small samples *Proc. Camb. Phil. Soc.*, Vol XXXII(1936), pp. 560–66.

[5] R. A. Fisher, "On a point raised by M. S. Bartlett on fiducial probability," *Ann. Eugen*, Vol. VIII(1937), pp. 370–75.

[6] "Student," "The probable error of a mean," *Biometrika*, Vol. VI(1908), pp. 1–25.

[7] P. V. Sukhatme, "On Fisher and Behrens' test of significance for the difference in means of two normal samples," *Sankhyā*, Vol. IV(1938), pp. 39–48

[8] R. A. Fisher, "Samples with possibly unequal variances," *Ann. Eugen*, Vol IX(1939), pp. 174–180.

[9] F. Yates, "An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters," *Proc. Camb. Phil. Soc.* (in press).

University College,
London.

A NOTE ON NEYMAN'S THEORY OF STATISTICAL ESTIMATION[1]

By Solomon Kullback

In this note we shall examine a section of a recent paper by Neyman[1] dealing with statistical estimation. Consider the following quotation from that section[2] which deals with the statement of the problem:

"Consider the variables $[x_1, x_2, \cdots, x_n]$ and assume that the form of the probability law $[p(x_1, \cdots, x_n \mid \theta_1, \theta_2, \cdots, \theta_l)]$ is known, that it involves the parameters $\theta_1, \theta_2, \cdots, \theta_l$ which are constant (not random variables), and that the numerical values of these parameters are unknown. It is desired to estimate one of these parameters, say θ_1. By this I shall mean that it is desired to define two functions $\bar{\theta}(E)$ and $\underline{\theta}(E) \leq \bar{\theta}(E)$, determined and single valued at any point E of the sample space, such that if E' is the sample point determined by observation, we can (1) calculate the corresponding values of $\underline{\theta}(E')$ and $\bar{\theta}(E')$ and (2), state that the true value of θ_1, say θ_1^0, is contained within the limits

$$\underline{\theta}(E') \leq \theta_1^0 \leq \bar{\theta}(E') \tag{18}$$

this statement having some intelligible justification on the ground of the theory of probability.

---

[1] Specifically we refer to J. Neyman "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil Trans. Roy. Soc*, vol. A236 (1937), pp 333–380.

[2] J. Neyman, loc. cit., p. 347. The material in brackets are slight alterations of the original text in order that the quotation do not refer to previous matter in the original paper.

This point requires to be made more precise. Following the routine of thought established under the influence of the Bayes Theorem, we could ask that, given the sample point E', the probability of θ_1^0, *falling within the limits* (18) should be large, say $\alpha = 0.99$, etc. If we expressed this condition by the formula

$$P\{\underline{\theta}(E') < \theta_1^0 < \bar{\theta}(E') \mid E'\} = \alpha \tag{19}$$

we see at once that it contradicts the assumption that θ_1^0 is constant. In fact, on this assumption, whatever the fixed point E' and the values $\underline{\theta}(E')$ and $\bar{\theta}(E')$, the only values the probability (19) may possess are zero and unity. For this reason we shall drop the specification of the problem as given by the condition (19)."

We believe that the following approach to the problem, emphasizes to a greater extent the fact that if the practical statistician follows the steps recommended as a result of Neyman's solution, then 'in the long run he will be correct in about 100α percent of all cases'.

Let us return again to the condition (19) of the quotation, and write

$$(1) \qquad \pi(E) = P\{\underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) \mid E\}$$

where of course $\pi(E)$ = zero or unity according as the true value of θ_1, say θ_1^0 does not or does satisfy the inequality

$$(2) \qquad \underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E)$$

We may however calculate the *average value* of $\pi(E)$ i.e., the percentage of cases in which in the long run the statistician will be correct.[3] In accordance with the definition of an average

$$(3) \qquad \overline{\pi(E)} = \int_R \pi(E) p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1\, dx_2 \cdots dx_n$$

where the region R is the entire sample space. If we let R_1 be that portion of the sample space for which (2) is satisfied, then since $\pi(E) = 1$ if E falls in R_1 and zero otherwise

$$(4) \qquad \overline{\pi(E)} = \int_{R_1} p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1\, dx_2 \cdots dx_n$$

Thus, if we want our rule to lead to a correct statement in 100α percent of cases in the long run, we must look for two functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ such that for the corresponding region R_1

$$(5) \qquad \overline{\pi(E)} = \int_{R_1} p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1\, dx_2 \cdots dx_n = \alpha$$

holds good whatever the value θ_1^0 of θ_1 and whatever the values of the other parameters $\theta_2, \theta_3, \cdots, \theta_l$ involved in the probability law of the X's may be.

[3] Cf. A. Werteimer, "A Note on Confidence Intervals and Inverse Probability," *Annals Math. Statistics*, Vol. X (1939), pp. 74ff.

If we apply to the preceding the calculus of probability in accordance with Neyman,[4] we find that (5) may be written as

$$\overline{\pi(E)} = P\{\underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) \mid \theta_1^0\} = \alpha \tag{6}$$

which, with the conditions stated for (5) is identical with formula (20) on page 348 of Neyman's paper.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, D. C.

[4] J. Neyman loc. cit. pp. 333-343.

A NOTE ON A PRIORI INFORMATION

BY C. EISENHART

A survey of recent literature on mathematical statistics is sufficient to reveal the fact that in approaching certain types of problems some writers assume more information known *a priori* than do other writers. Indeed, it soon becomes evident that great care is necessary in wording (and in reading) propositions in mathematical statistics. Furthermore, propositions which are true and powerful when certain information is known *a priori* may become either useless or irrelevant according as more, or less, information is available *a priori*. Once this situation is appreciated some apparent contradictions are resolved, and certain exceptional examples can "be reasonably regarded as bearing out the principle to which formally they are anomalous."

So far as I know it was Bartlett [1, p. 271] who first clearly pointed out how a slight change in the amount of information known *a priori* can greatly alter the complexion of a problem. He was indebted to Neyman and Pearson [5, p. 122] for his problem, which was to develop a test of the statistical hypothesis, H_0, that $\beta = \beta_0$ *and* $\gamma = \gamma_0$ for a random sample from the distribution

$$p(x) = \begin{cases} \beta e^{-\beta(x-\gamma)} & \text{for } x \geq \gamma \\ 0 & \text{for } x < \gamma. \end{cases} \tag{1}$$

If (1) expresses *all* the information (about the distribution of x) that is to be considered as known *a priori*, any value of $\beta > 0$ and any finite value of γ being admissible, then it follows immediately from a result of R. A. Fisher's [2, p. 295] that no uniformly most powerful test, in the sense of Neyman and Pearson [4; 5, p. 115], can exist for H_0, since H_0 involves the simultaneous testing of two unrelated parameters.[1]

[1] Since Fisher's wording is important it will be well to quote him here: "It is evident, at once, that such a system [of maximum likelihood relations needed to insure the existence of a uniformly most powerful test] is only possible when the class of hypotheses considered

By assuming that in addition to (1) the *a priori information includes* the knowledge that $\beta \geq \beta_0$ and $\gamma \leq \gamma_0$ constitute the only admissible ranges of values of these parameters, Neyman and Pearson [5, p. 122] have succeeded in showing that a uniformly most powerful test of H_0 does exist when the admissible values of β and γ are restricted in this way. At first this appears to be in contradiction to Fisher's statement referred to above, but Bartlett [1, p. 271] points out that the restrictions on the admissible values of β and γ reduce the problem *effectively* to one of testing a single parameter: In the first place, no statistical test is necessary if an observation less than γ_0 occurs, since this refutes the hypothesis H_0 immediately. Therefore, a statistical test of H_0 is needed only when none of the observations are less than γ_0, and *for such observations* the distribution law is

$$p(x) = \beta e^{-\beta(x-\gamma)}/e^{-\beta(\gamma_0-\gamma)} = \beta e^{-\beta(x-\gamma_0)}, \quad x \geq \gamma_0, \tag{2}$$

and is independent of γ. In consequence, the test reduces to testing the single parameter β in (2), for which the arithmetic mean, $\bar{x}$, is a sufficient statistic. The discovery of a uniformly most powerful test of H_0, when the above restrictions are placed on the admissible values of β and γ, is, therefore, reasonably consistent with the full meaning of Fisher's statement.

The preceding example makes quite clear how a little additional *a priori* knowledge can affect the solution of a problem in mathematical statistics. The *a priori* knowledge employed by writers in mathematical statistics usually falls into one of the following categories:

(i) The elementary probability law is taken to be continuous or discrete, as the case may be, but its mathematical form is left unspecified.

(ii) The elementary probability law is taken to be of a definite mathematical form involving one or more parameters the value(s) of which is (are) not considered as known *a priori*, and any value(s) of this (these) parameter(s) consistent with the non-negative character of a probability law is (are) admissible.

(iii) Here the information assumed known is as in (ii) except that the admissible values of the parameter(s) form (a) restricted sub-set (or sub-sets) of the values admissible in (ii), such subsets, however, being comprised of more than a single value.

(iv) The information is so complete that the admissible values of the parameter(s) have (a) known *a priori* probability distribution(s)—if a param-

---

involves only a single parameter θ, or, what comes to the same thing, when all the parameters entering into the specification of the population are definite functions of one of their number. In this case, the regions defined by the uniformly most powerful test of significance are those defined by the estimate of the maximum likelihood, T. For the test to be uniformly most powerful, moreover, these regions must be independent of θ, showing that the statistic must be of the special type distinguished as sufficient." (Words in square brackets are mine.—C. E.)

eter θ is known to have a definite value θ', then the *a priori* probability law of θ can be taken as (Prob. θ equals θ') = 1, (Prob. θ not equal to θ') = 0.[2]

As statistical theory advances it may become necessary to classify problems according to the amount of information which may be assumed known *a priori*, when proceeding to their solution. No claim is made here that the above categories are the best to choose, but it may prove fruitful to study the extent to which results obtained with a certain amount of information assumed known are useful when more, less, or perhaps different, information is taken as known *a priori* In particular, as the preceding example shows, it may be well to investigate exactly what are the implications of restricting the ranges of the admissible values of parameters.

It is unwise to attempt to predict the outcome of such research at this time, but it is probably safe to say that an increase in *a priori* information will generally render possible better tests of significance—better in the sense that, for a given probability of rejecting the hypothesis tested when true, the probability of rejecting it when false will be greater—and narrower confidence intervals for a given confidence level. The example already given, concerned with a test of significance, supports this conjecture. As a further example, from the point of view of estimation, we may recall that it is possible with a level of confidence equal to .96875 to assert [3, p. 4] that the true *median* of the population from which a random sample of 6 was drawn lies within the observed range of the sample, and this without any assumption about the population except that it is continuous. If, however, the population is known to be of normal form with unknown mean, m, and standard deviation, σ, then Student's t will provide the narrowest confidence intervals for the median of the population, since t provides [6, p 378] the best available confidence intervals for the mean, m, (which is also the median) of a normal population when σ is unknown—if the population is normal and σ is known, then the normal deviate $(\bar{x} - m)\sqrt{6}/\sigma$ will supply still narrower confidence limits for m.

In conclusion, the circumstances under which it may be desired to apply methods of statistical inference may differ considerably in the amount of knowledge available to the research worker *a priori*, and the most efficient tests of significance and methods of estimation applicable to a given case will depend upon the nature of the available information as described in the above classification. In comparing the procedures of different writers, therfore, it is most important to examine their premises and see how much information each is considering as known at the start.

REFERENCES

[1] M. S. Bartlett, "Properties of Sufficiency and Statistical Tests" *Proc. Royal Society of London*, Series A, vol. 160(1937), pp. 268-282.

[2] Category (iv) should be regarded as sharing certain 'territory' with the others since, for instance, a problem may consist in the estimation of the true mean, m, which has a known *a priori* distribution, without assuming the mathematical form of the probability law of the random variable observed, x, as known *a priori*.

[2] R. A Fisher, "Two New Properties of Mathematical Likelihood" *Proc. Royal Society of London*, Series A, vol. 144(1934), pp. 285-307.
[3] R. A Fisher, " 'Student' " *Annals of Eugenics*, vol. 9(1939) pp. 1-9
[4] J. Neyman and E S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Phil. Trans Royal Society of London*, Series A, vol 231(1933), pp. 289-337.
[5] J Neyman and E. S. Pearson, "Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses." *Statistical Research Memoirs*, vol. 1(1936), pp. 113-137.
[6] J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." *Phil. Trans. Royal Society of London*, Series A, vol. 236(1937), pp. 333-380.

University of Wisconsin
Madison, Wisconsin

A NOTE ON COMPUTATION FOR ANALYSIS OF VARIANCE

By Morris C. Bishop

The method of computation for analysis of variance commonly favored is one which involves obtaining the total and total sum of squares in a single operation on a computing or card-punch machine,[1] in which case a check on the accuracy of the work requires complete recomputation. But the best tools available to the student, and sometimes to the experimenter, are a table of squares and perhaps a listing machine. In such a situation, a simple algorithm which embodies checks on the computations is urgently needed. The method here presented reduces the arithmetic to repeated application of a single procedure, with adequate checks; it reveals rather than obscures the sample variances, which may or may not be of primary importance; and it provides an intuitively logical portrayal of the step-by-step improvement of the estimate of population variance.

The data items and their squares may be merged into a single table by setting them down in staggered fashion, as shown in Table I. If only a single criterion of classification is to be used—classified into columns, say—the columns are summed down, and then these totals across (obtained as two sets of subtotals and totals on a listing machine). This yields the grand total (T) and total sum of squares $\left(\sum_{i=1}^{N}\sum_{j=1}^{k} X_{ij}^2\right)$. Summing across and down verifies the addition and provides material for two-way classification. The total sum of squares of deviations is obtained by the familiar formula

$$(1) \qquad \sum_{i=1}^{N}\sum_{j=1}^{k} (X_{ij} - \bar{X})^2 = \sum_{i=1}^{N}\sum_{j=1}^{k} X_{ij}^2 - \frac{T^2}{Nk}$$

where Nk is the total number of observations in N rows and k columns.

[1] See George W. Snedecor, *Analysis of Variance and Covariance*, and Paul R. Rider, *Modern Statistical Methods*.